CDC NSSP ESSENCE
In-person Training Workshop
Student Packet

Content was developed for and funded by the Centers for Disease Control and Prevention (CDC) for training purposes. The findings and conclusions in this presentation are those of the authors and do not necessarily represent the views of CDC.

Center for Surveillance, Epidemiology, and Laboratory Services
Division of Health Informatics and Surveillance

ESSENCE Overview

ESSENCE Training Workshop

ESSENCE Overview

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

ESSENCE & NSSP

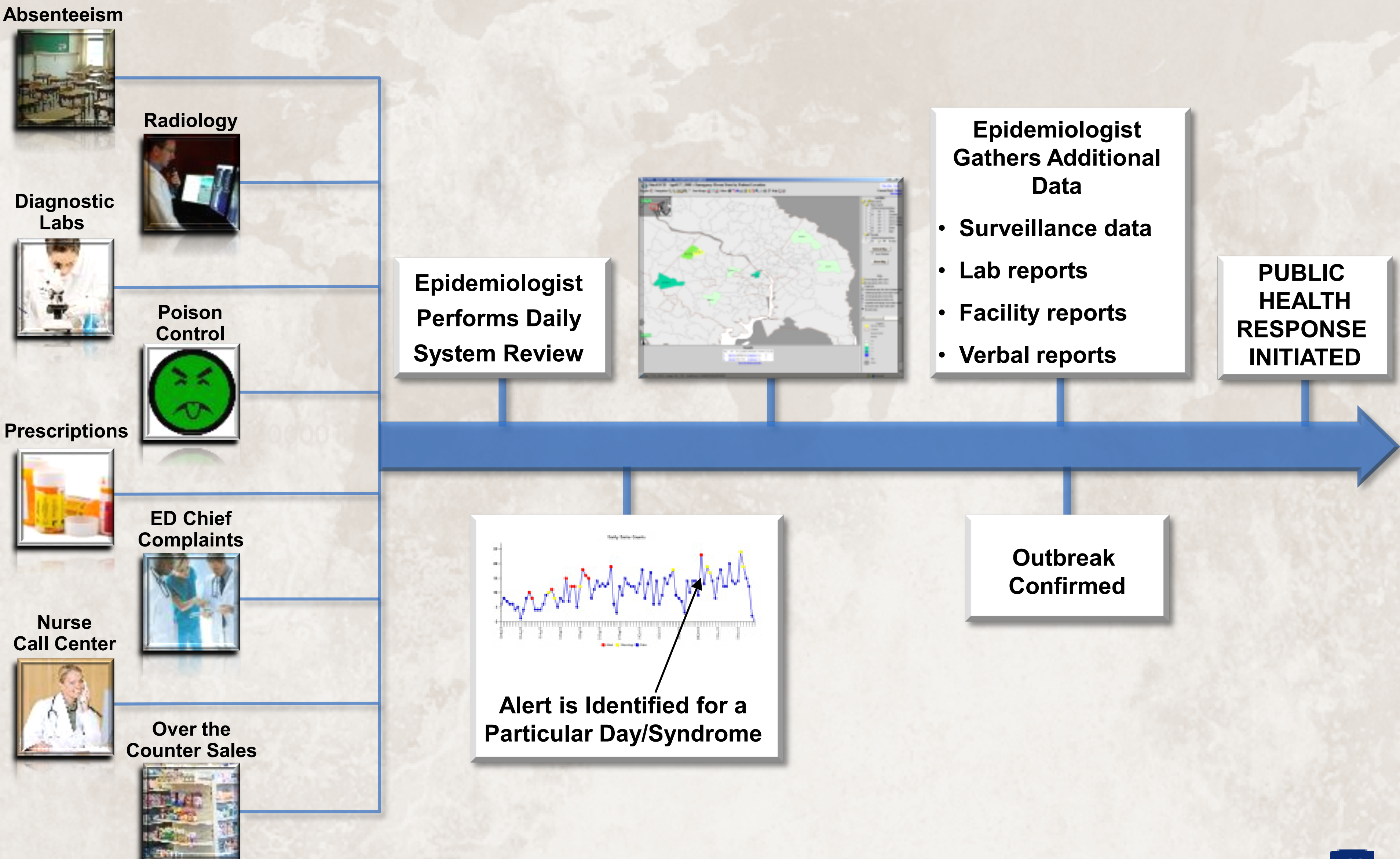
- **ESSENCE was identified as a new tool available on the NSSP Platform**
- **Goal is to help CDC improve data quality, efficiency, and usefulness of data collected as part of the NSSP**

What is ESSENCE ?

E lectronic
S urveillance
S ystem for the
E arly
N otification of
C ommunity-based
E pidemics

Web-based disease surveillance information system developed to alert Health Authorities of infectious disease outbreaks, including possible bioterrorism attacks

Electronic Disease Surveillance

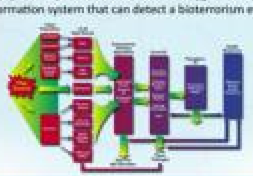


The Evolution of ESSENCE

Wayne Loschen, Sheryl Happel Lewis, Richard Wojcik, Howard Burkom, Joseph Lombardo
The Johns Hopkins University Applied Physics Laboratory (JHU/APL)

97 98 99 00 01 02 03 04 05 06 07 08 09 10 11 12 13 14

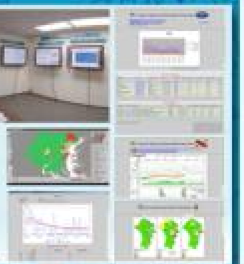
Early research in information management leads to a request by Dr. Georges Benjamin and collaboration with the Maryland Dept. of Health and Mental Hygiene to develop an information system that can detect a bioterrorism event.



Y2K: The precursor to the current ESSENCE system, the Maryland Disease Surveillance and Reporting System (MDSRS), goes live for Y2K surveillance. This leads to the DARPA Bio-ALIRT program in which JHU/APL teams with DoD-GEIS to improve detector algorithms and the early system architecture.

ESSENCE moves from a dynamic web framework to a static pre-built one, allowing use of GIS technologies from ESRI (ArcView) to produce maps for users. However, the lack of users' capability to make changes forces a rapid return to a dynamic web framework.

September 12th, 2001: This afternoon, ESSENCE II goes from being a research project to a live operational system in Maryland. ESSENCE has been converted from a static web framework to a fully dynamic version with a database backend.



Iraq War: In response to the need to monitor deployed troops, ESSENCE III is developed and deployed in March 2003. In 3 weeks, the ESSENCE team builds an entirely new version of ESSENCE with new datasources, customizable groupings, and the first incorporation of an outside detector. This substantiates the value of adding an Application Program Interface (API) to the detection system so new algorithms can be plugged into ESSENCE.

SARS: The need to begin looking for a very specific case definition instead of a general syndrome is supported by the decision to allow free-text chief-complaint queries into the system. This event brings awareness to the power of these types of tools as information systems—not just detection systems. It is also the last time "syndromic surveillance" is used universally. After this date, "Disease Surveillance" becomes the description.

Expansion: The DoD JSIPP and the DHS BioWatch proliferate ESSENCE instances across the country. This influx of new users and perspectives leads to a new focus on enhancing visualization flexibility and exportability. Users explain that the need to share results with peers to convince them of the necessity for action is just as important as finding an outbreak in the first place. Features are added to support better labeling and download capabilities.

ANCR: In 2004, Maryland, Virginia, and the District of Columbia agree to share data across state lines, and the "Aggregate National Capital Region" version of ESSENCE is created. This sparks the need for the "Enhanced Surveillance Operating Group (ESOG)," a group of users who help guide future development efforts in ESSENCE. This close tie with users becomes a foundation for the ESSENCE system.

2007 Super Bowl & 2009 Presidential Inauguration: These high-profile events provide the opportunity to test information-sharing strategies and deploy the InfoShare tool to allow NCR ESSENCE users the ability to share information with Federal partners during the event.

Real-Life Algorithms: "Stuff happens," especially to networks' and data providers' ability to send data consistently. The ESSENCE team recognizes that detection algorithms need to adapt to the real-world situation of missing data, outliers, and overall erratic behaviors. In response to this, ESSENCE algorithms are enhanced with Outlier Removal, Zero Count Filtration, and a new switching system to move from linear regression to EWMA to Poisson as needed. Working with the state of Florida, the team develops a Time-of-Arrival (TOA)-based detection system to look for small clusters of abnormal patients.

Open Source: Working with DoD-GEIS, the ESSENCE team identifies the need to transition the technology and lessons learned from ESSENCE to open-source versions that are easy to operate, free, and that can be deployed around the world. The SAGES toolkit is developed, including OpenESSENCE and ESSENCE: Desktop Edition.

Full EMR: Working with the Veterans Affairs and Dept. of Defense, the ESSENCE team investigates the benefits of and issues with including elements of an EMR beyond the demographics and chief complaints. New algorithms that detect the severity of patients to better characterize an outbreak are built. Working with multi-terabyte database sizes prepares the ESSENCE system for dealing with large datasets in every jurisdiction.

After the Boston bombings, ESSENCE was adapted to look for non-infectious health events including: Anxiety, Depression, Suicidal Tendencies, and Hearing Loss. The system helped track the health of the Boston communities as they dealt with the aftermath of the attacks.

BOSTON PUBLIC HEALTH COMMISSION

The Centers for Disease Control and Prevention announced that the National Syndromic Surveillance Program (NSSP) is transitioning to new syndromic surveillance software—the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE)—which will be hosted on the cloud-based BioSense Platform.

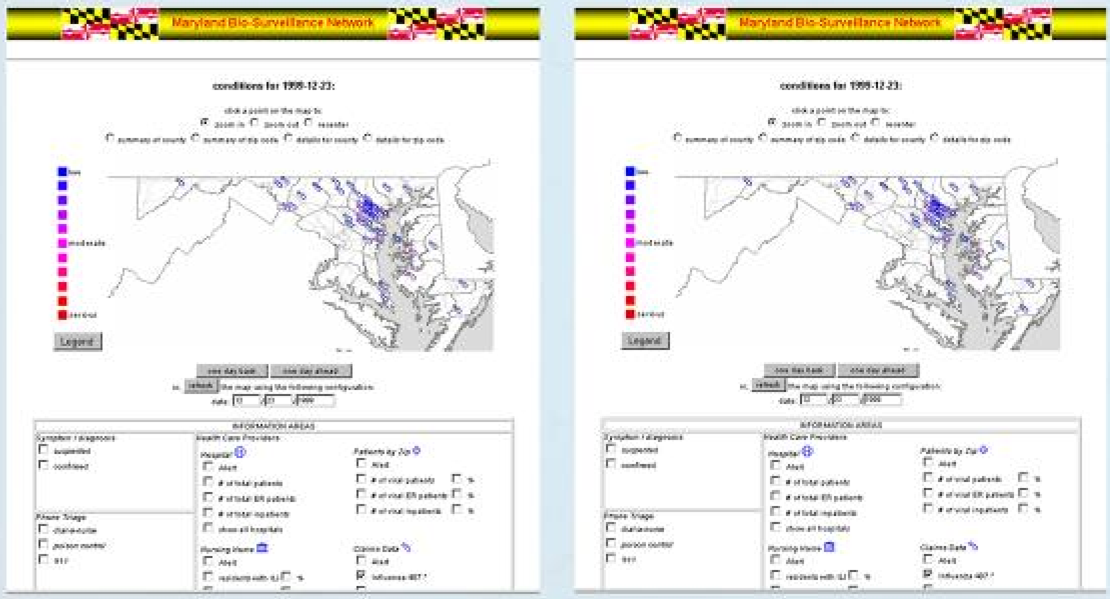
CDC CENTERS FOR DISEASE CONTROL AND PREVENTION

NSSP National Syndromic Surveillance Program BioSense Platform

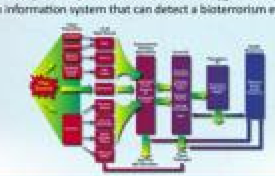
The Cloud version of ESSENCE

ESSENCE could not have been built without the support of many sponsors, numerous collaborators, perceptive users, and all of the dedicated members of the ESSENCE team.

Y2K: The precursor to the current ESSENCE system, the Maryland Disease Surveillance and Reporting System (MDSRS), goes live for Y2K surveillance. This leads to the DARPA Bio-ALIRT program in which JHU/APL teams with DoD-GEIS to improve detector algorithms and the early system architecture.



Early research in information management leads to a request by Dr. Georges Benjamin and collaboration with the Maryland Dept. of Health and Mental Hygiene to develop an information system that can detect a bioterrorism event.



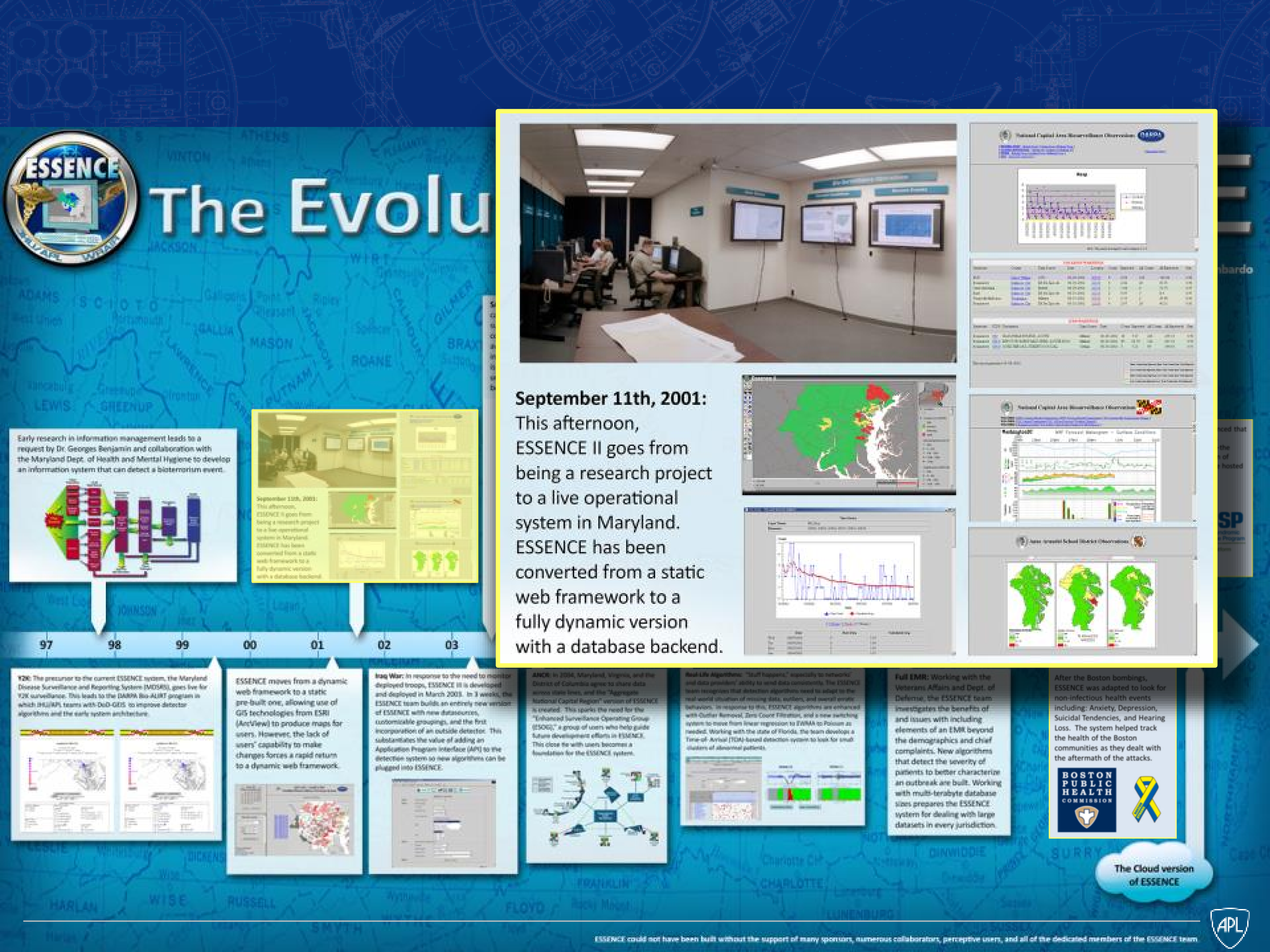
97 98 99 … 12 13 14

The Centers for Disease Control and Prevention announced that the National Syndromic Surveillance Program (NSSP) is transitioning to new syndromic surveillance software—the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE)—which will be hosted on the cloud-based BioSense Platform.

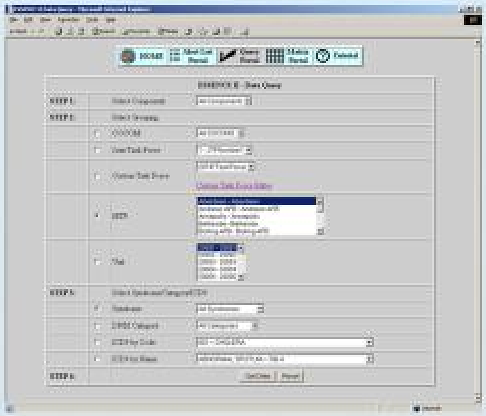
After the Boston bombings, ESSENCE was adapted to look for non-infectious health events including: Anxiety, Depression, Suicidal Tendencies, and Hearing Loss. The system helped track the health of the Boston communities as they dealt with the aftermath of the attacks.

The Cloud version of ESSENCE

ESSENCE could not have been built without the support of many sponsors, numerous collaborators, perceptive users, and all of the dedicated members of the ESSENCE team.


ESSENCE
The Evolu
September 11th, 2001: This afternoon, ESSENCE II goes from being a research project to a live operational system in Maryland. ESSENCE has been converted from a static web framework to a fully dynamic version with a database backend.
National Capital Area Biosurveillance Observations
Anne Arundel School District Observations
Early research in information management leads to a request by Dr. Georges Benjamin and collaboration with the Maryland Dept. of Health and Mental Hygiene to develop an information system that can detect a bioterrorism event.
97
98
99
00
01
02
03
Y2K: The precursor to the current ESSENCE system, the Maryland Disease Surveillance and Reporting System (MDSRS), goes live for Y2K surveillance. This leads to the DARPA Bio-ALIRT program in which JHU/APL teams with DoD-GEIS to improve detector algorithms and the early system architecture.
ESSENCE moves from a dynamic web framework to a static pre-built one, allowing use of GIS technologies from ESRI (ArcView) to produce maps for users. However, the lack of users' capability to make changes forces a rapid return to a dynamic web framework.
Iraq War: In response to the need to monitor deployed troops, ESSENCE III is developed and deployed in March 2003. In 3 weeks, the ESSENCE team builds an entirely new version of ESSENCE with new datasources, customizable groupings, and the first incorporation of an outside detector. This substantiates the value of adding an Application Program Interface (API) to the detection system so new algorithms can be plugged into ESSENCE.
ANCR: In 2004, Maryland, Virginia, and the District of Columbia agree to share data across state lines, and the "Aggregate National Capital Region" version of ESSENCE is created. This sparks the need for the "Enhanced Surveillance Operating Group (ESOG)," a group of users who help guide future development efforts in ESSENCE. This close tie with users becomes a foundation for the ESSENCE system.
Real-Life Algorithms: "Stuff happens," especially to networks' and data providers' ability to send data consistently. The ESSENCE team recognizes that detection algorithms need to adapt to the real world situation of missing data, outliers, and overall erratic behaviors. In response to this, ESSENCE algorithms are enhanced with Outlier Removal, Zero Count Filtration, and a new switching system to move from linear regression to EWMA to Poisson as needed. Working with the state of Florida, the team develops a Time-of-Arrival (TOA)-based detection system to look for small clusters of abnormal patients.
Full EMR: Working with the Veterans Affairs and Dept. of Defense, the ESSENCE team investigates the benefits of and issues with including elements of an EMR beyond the demographics and chief complaints. New algorithms that detect the severity of patients to better characterize an outbreak are built. Working with multi-terabyte database sizes prepares the ESSENCE system for dealing with large datasets in every jurisdiction.
After the Boston bombings, ESSENCE was adapted to look for non-infectious health events including: Anxiety, Depression, Suicidal Tendencies, and Hearing Loss. The system helped track the health of the Boston communities as they dealt with the aftermath of the attacks.
BOSTON PUBLIC HEALTH COMMISSION
The Cloud version of ESSENCE
ESSENCE could not have been built without the support of many sponsors, numerous collaborators, perceptive users, and all of the dedicated members of the ESSENCE team.
APL

Iraq War: In response to the need to monitor deployed troops, ESSENCE III is developed and deployed in March 2003. In 3 weeks, the ESSENCE team builds an entirely new version of ESSENCE with new datasources, customizable groupings, and the first incorporation of an outside detector. This substantiates the value of adding an Application Program Interface (API) to the detection system so new algorithms can be plugged into ESSENCE.



tion of ESSENCE

Wayne Loschen, Sheryl Happel Lewis, Richard Wojcik, Howard Burkom, Joseph Lombardo
The Johns Hopkins University Applied Physics Laboratory (JHU/APL)

SARS: The need to begin looking for a very specific case definition instead of a general syndrome is supported by the decision to allow free-text chief-complaint queries into the system. This event brings awareness to the power of these types of tools as information systems—not just detection systems. It is also the last time "syndromic surveillance" is used universally. After this date, "Disease Surveillance" becomes the description.

Expansion: The DoD JSIPP and the DHS BioWatch proliferate ESSENCE instances across the country. This influx of new users and perspectives leads to a new focus on enhancing visualization flexibility and exportability. Users explain that the need to share results with peers to convince them of the necessity for action is just as important as finding an outbreak in the first place. Features are added to support better labeling and download capabilities.

2007 Super Bowl & 2009 Presidential Inauguration: These high-profile events provide the opportunity to test information-sharing strategies and deploy the InfoShare tool to allow NCR ESSENCE users the ability to share information with Federal partners during the event.

Open Source: Working with DoD-GEIS, the ESSENCE team identifies the need to transition the technology and lessons learned from ESSENCE to open-source versions that are easy to operate, free, and that can be deployed around the world. The SAGES toolkit is developed, including OpenESSENCE and ESSENCE Desktop Edition.

The Centers for Disease Control and Prevention announced that the National Syndromic Surveillance Program (NSSP) is transitioning to new syndromic surveillance software—the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE)—which will be hosted on the cloud-based BioSense Platform.

04 05 06 07 08 09 10 11 12 13 14

Y2K: The precursor to the current ESSENCE system, the Maryland Disease Surveillance and Reporting System (MDSRS), goes live for Y2K surveillance. This leads to the DARPA Bio-ALIRT program in which JHU/APL teams with DoD-GEIS to improve detector algorithms and the early system architecture.

ESSENCE moves from a dynamic web framework to a static pre-built one, allowing use of GIS technologies from ESRI (ArcView) to produce maps for users. However, the lack of users' capability to make changes forces a rapid return to a dynamic web framework.

Iraq War: In response to the need to monitor deployed troops, ESSENCE III is developed and deployed in March 2003. In 3 weeks, the ESSENCE team builds an entirely new version of ESSENCE with new datasources, customizable groupings, and the first incorporation of an outside detector. This substantiates the value of adding an Application Program Interface (API) to the detection system so new algorithms can be plugged into ESSENCE.

ANCR: In 2004, Maryland, Virginia, and the District of Columbia agree to share data across state lines, and the "Aggregate National Capital Region" version of ESSENCE is created. This sparks the need for the "Enhanced Surveillance Operating Group (ESOG)," a group of users who help guide future development efforts in ESSENCE. This close tie with users becomes a foundation for the ESSENCE system.

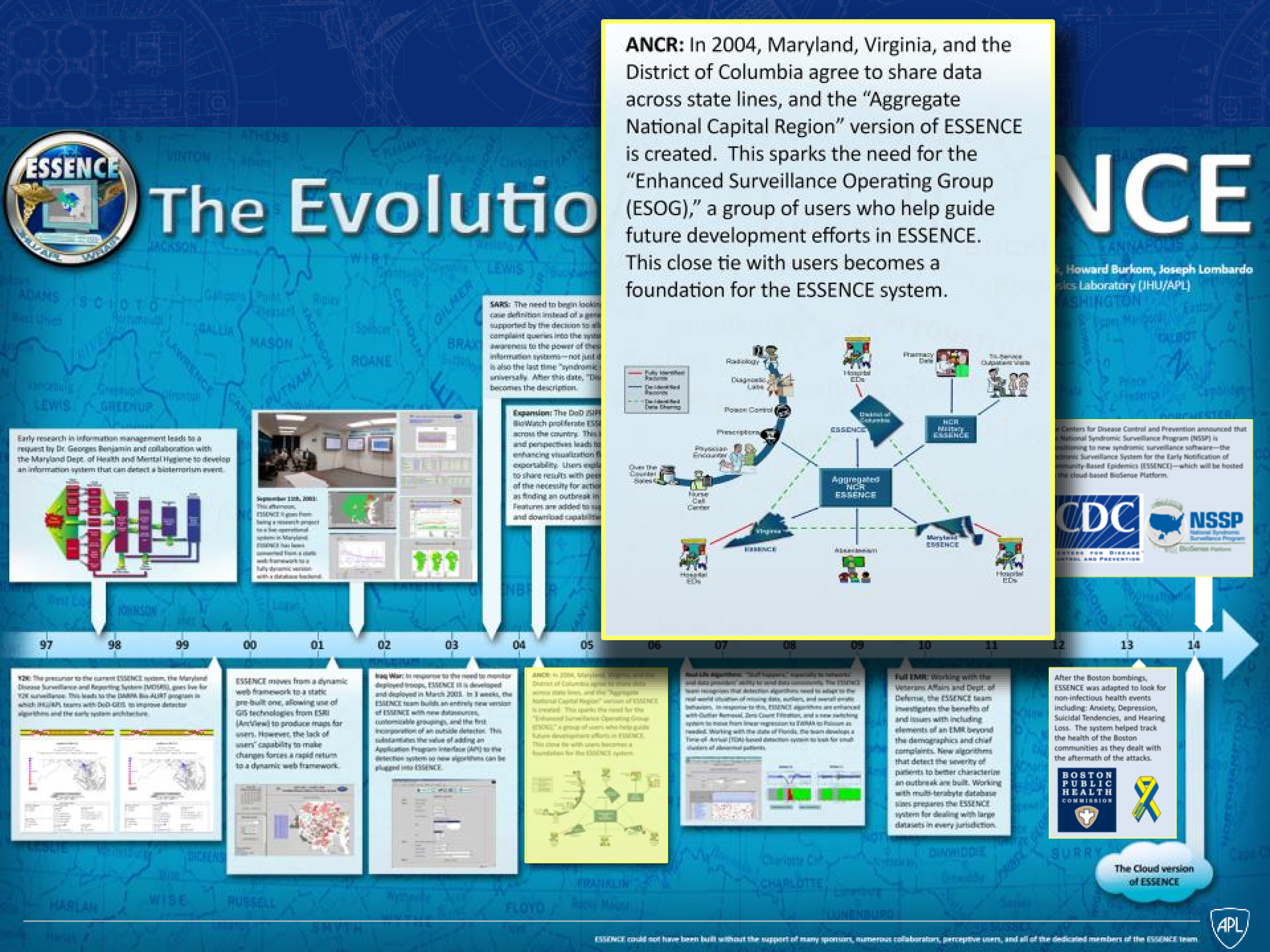
Real-Life Algorithms: "Stuff happens," especially to networks' and data providers' ability to send data consistently. The ESSENCE team recognizes that detection algorithms need to adapt to the real-world situation of missing data, outliers, and overall erratic behaviors. In response to this, ESSENCE algorithms are enhanced with Outlier Removal, Zero Count Filtration, and a new switching system to move from linear regression to EWMA to Poisson as needed. Working with the state of Florida, the team develops a Time-of-Arrival (TOA)-based detection system to look for small clusters of abnormal patients.

Full EMR: Working with the Veterans Affairs and Dept. of Defense, the ESSENCE team investigates the benefits of and issues with including elements of an EMR beyond the demographics and chief complaints. New algorithms that detect the severity of patients to better characterize an outbreak are built. Working with multi-terabyte database sizes prepares the ESSENCE system for dealing with large datasets in every jurisdiction.

After the Boston bombings, ESSENCE was adapted to look for non-infectious health events including: Anxiety, Depression, Suicidal Tendencies, and Hearing Loss. The system helped track the health of the Boston communities as they dealt with the aftermath of the attacks.

The Cloud version of ESSENCE

ESSENCE could not have been built without the support of many sponsors, numerous collaborators, perceptive users, and all of the dedicated members of the ESSENCE team.

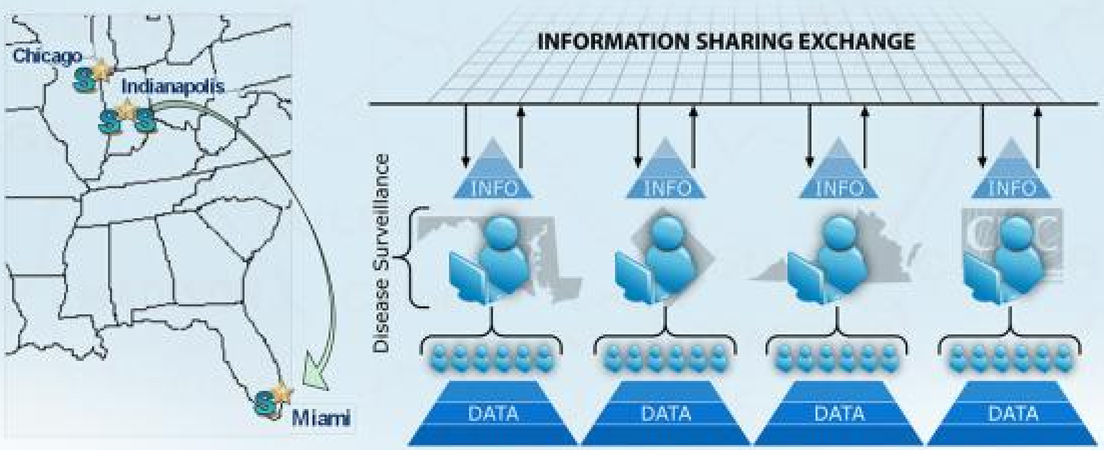


ESSENCE
The Evolution of ESSENCE
Howard Burkom, Joseph Lombardo
ANCR: In 2004, Maryland, Virginia, and the District of Columbia agree to share data across state lines, and the "Aggregate National Capital Region" version of ESSENCE is created. This sparks the need for the "Enhanced Surveillance Operating Group (ESOG)," a group of users who help guide future development efforts in ESSENCE. This close tie with users becomes a foundation for the ESSENCE system.
Radiology
Diagnostic Labs
Poison Control
Prescriptions
Physician Encounter
Over the Counter Sales
Nurse Call Center
Hospital EDs
District of Columbia ESSENCE
Pharmacy Data
Tri-Service Outpatient Visits
NCR Military ESSENCE
Aggregated NCR ESSENCE
Virginia ESSENCE
Absenteeism
Maryland ESSENCE
Hospital EDs
Early research in information management leads to a request by Dr. Georges Benjamin and collaboration with the Maryland Dept. of Health and Mental Hygiene to develop an information system that can detect a bioterrorism event.
97
98
99
00
01
02
03
04
05
06
07
08
09
10
11
12
13
14
ESSENCE moves from a dynamic web framework to a static pre-built one, allowing use of GIS technologies from ESRI (ArcView) to produce maps for users. However, the lack of users' capability to make changes forces a rapid return to a dynamic web framework.
Iraq War: In response to the need to monitor deployed troops, ESSENCE III is developed and deployed in March 2003. In 3 weeks, the ESSENCE team builds an entirely new version of ESSENCE with new datasources, customizable groupings, and the first incorporation of an outside detector. This substantiates the value of adding an Application Program Interface (API) to the detection system so new algorithms can be plugged into ESSENCE.
Full EMR: Working with the Veterans Affairs and Dept. of Defense, the ESSENCE team investigates the benefits of and issues with including elements of an EMR beyond the demographics and chief complaints. New algorithms that detect the severity of patients to better characterize an outbreak are built. Working with multi-terabyte database sizes prepares the ESSENCE system for dealing with large datasets in every jurisdiction.
After the Boston bombings, ESSENCE was adapted to look for non-infectious health events including: Anxiety, Depression, Suicidal Tendencies, and Hearing Loss. The system helped track the health of the Boston communities as they dealt with the aftermath of the attacks.
BOSTON PUBLIC HEALTH COMMISSION
CDC
NSSP
The Cloud version of ESSENCE
ESSENCE could not have been built without the support of many sponsors, numerous collaborators, perceptive users, and all of the dedicated members of the ESSENCE team.
APL

The Evolution of ESSENCE

Wayne Loschen, Sheryl Happel Lewis, Richard Wojcik, Howard Burkom, Joseph Lombardo
The Johns Hopkins University Applied Physics Laboratory (JHU/APL)

2007 Super Bowl & 2009 Presidential Inauguration:
These high-profile events provide the opportunity to test information-sharing strategies and deploy the InfoShare tool to allow NCR ESSENCE users the ability to share information with Federal partners during the event.



Open Source: Working with DoD-GEIS, the ESSENCE team identifies the need to transition the technology and lessons learned from ESSENCE to open-source versions that are easy to operate, free, and that can be deployed around the world. The SAGES toolkit is developed, including OpenESSENCE and ESSENCE: Desktop Edition.

The Centers for Disease Control and Prevention announced that the National Syndromic Surveillance Program (NSSP) is transitioning to new syndromic surveillance software—the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE)—which will be hosted on the cloud-based BioSense Platform.

09 10 11 12 13 14

Full EMR: Working with the Veterans Affairs and Dept. of Defense, the ESSENCE team investigates the benefits of and issues with including elements of an EMR beyond the demographics and chief complaints. New algorithms that detect the severity of patients to better characterize an outbreak are built. Working with multi-terabyte database sizes prepares the ESSENCE system for dealing with large datasets in every jurisdiction.

After the Boston bombings, ESSENCE was adapted to look for non-infectious health events including: Anxiety, Depression, Suicidal Tendencies, and Hearing Loss. The system helped track the health of the Boston communities as they dealt with the aftermath of the attacks.

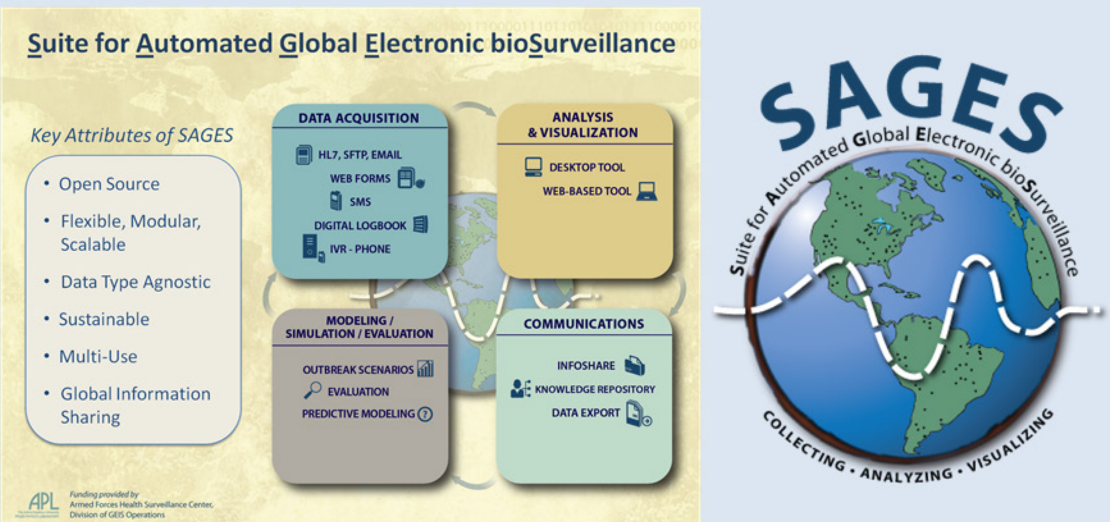
The Cloud version of ESSENCE

ESSENCE could not have been built without the support of many sponsors, numerous collaborators, perceptive users, and all of the dedicated members of the ESSENCE team.

The Evolution of ESSENCE

Wayne Loschen, Sheryl Happel Lewis, Richard Wojcik, Howard Burkom, Joseph Lombardo
The Johns Hopkins University Applied Physics Laboratory (JHU/APL)

Working with DoD-GEIS, the ESSENCE team identifies the need to transition the technology and lessons learned from ESSENCE to open source versions that are easy to operate, free, and can be deployed around the world. The SAGES toolkit is developed; including OpenESSENCE and ESSENCE: Desktop Edition.



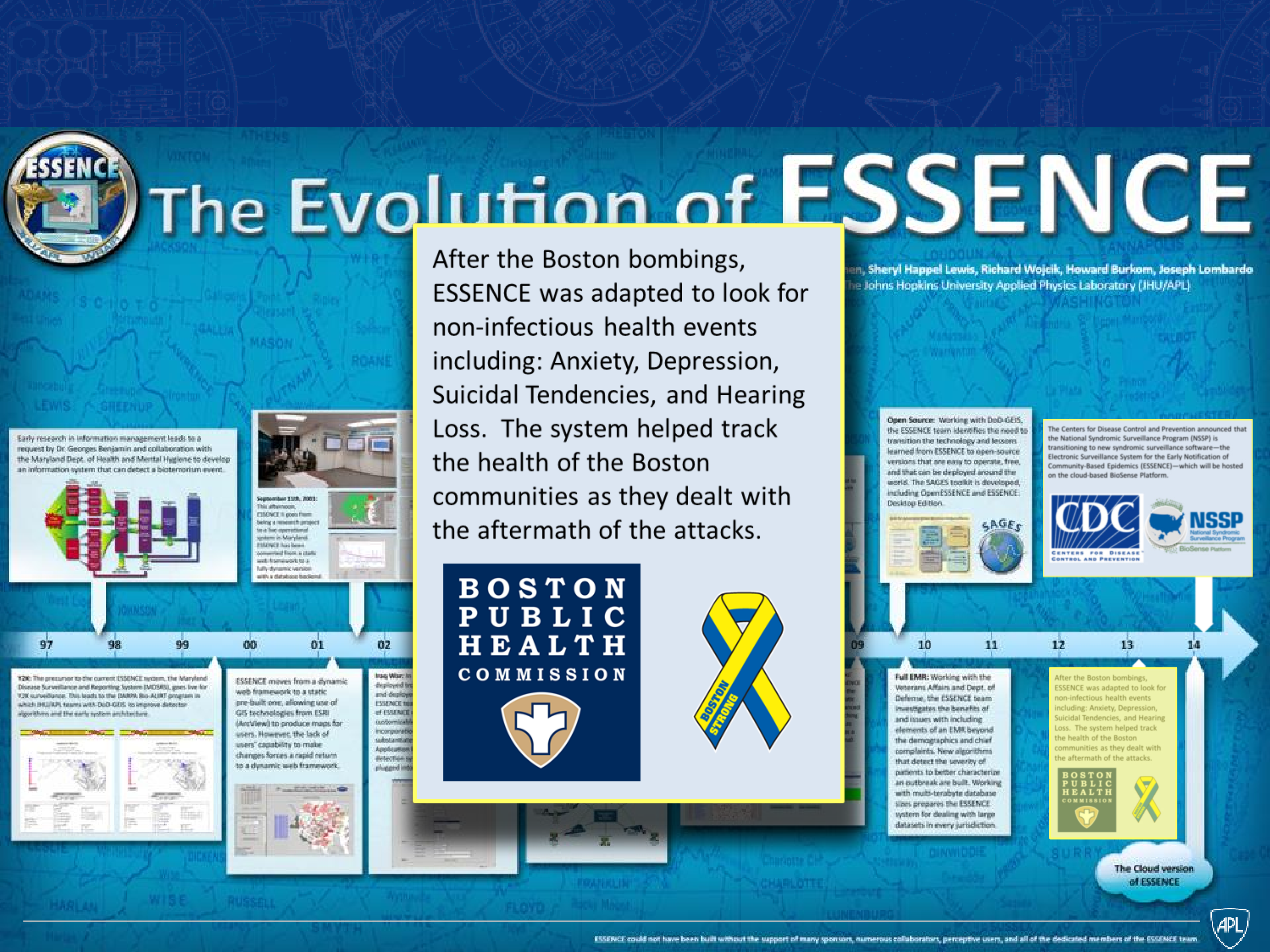
The Centers for Disease Control and Prevention announced that the National Syndromic Surveillance Program (NSSP) is transitioning to new syndromic surveillance software—the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE)—which will be hosted on the cloud-based BioSense Platform.

Full EMR: Working with the Veterans Affairs and Dept. of Defense, the ESSENCE team investigates the benefits of and issues with including elements of an EMR beyond the demographics and chief complaints. New algorithms that detect the severity of patients to better characterize an outbreak are built. Working with multi-terabyte database sizes prepares the ESSENCE system for dealing with large datasets in every jurisdiction.

After the Boston bombings, ESSENCE was adapted to look for non-infectious health events including: Anxiety, Depression, Suicidal Tendencies, and Hearing Loss. The system helped track the health of the Boston communities as they dealt with the aftermath of the attacks.

The Cloud version of ESSENCE

ESSENCE could not have been built without the support of many sponsors, numerous collaborators, perceptive users, and all of the dedicated members of the ESSENCE team.


The Evolution of ESSENCE
After the Boston bombings, ESSENCE was adapted to look for non-infectious health events including: Anxiety, Depression, Suicidal Tendencies, and Hearing Loss. The system helped track the health of the Boston communities as they dealt with the aftermath of the attacks.
BOSTON PUBLIC HEALTH COMMISSION
BOSTON STRONG
…en, Sheryl Happel Lewis, Richard Wojcik, Howard Burkom, Joseph Lombardo
The Johns Hopkins University Applied Physics Laboratory (JHU/APL)
Early research in information management leads to a request by Dr. Georges Benjamin and collaboration with the Maryland Dept. of Health and Mental Hygiene to develop an information system that can detect a bioterrorism event.
September 11th, 2001: This afternoon, ESSENCE II goes from being a research project to a live operational system in Maryland. ESSENCE has been converted from a static web framework to a fully dynamic version with a database backend.
Open Source: Working with DoD-GEIS, the ESSENCE team identifies the need to transition the technology and lessons learned from ESSENCE to open-source versions that are easy to operate, free, and that can be deployed around the world. The SAGES toolkit is developed, including OpenESSENCE and ESSENCE: Desktop Edition.
SAGES
The Centers for Disease Control and Prevention announced that the National Syndromic Surveillance Program (NSSP) is transitioning to new syndromic surveillance software—the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE)—which will be hosted on the cloud-based BioSense Platform.
CDC
NSSP
National Syndromic Surveillance Program
BioSense Platform
97
98
99
00
01
02
09
10
11
12
13
14
Y2K: The precursor to the current ESSENCE system, the Maryland Disease Surveillance and Reporting System (MDSRS), goes live for Y2K surveillance. This leads to the DARPA Bio-ALIRT program in which JHU/APL teams with DoD-GEIS to improve detector algorithms and the early system architecture.
ESSENCE moves from a dynamic web framework to a static pre-built one, allowing use of GIS technologies from ESRI (ArcView) to produce maps for users. However, the lack of users' capability to make changes forces a rapid return to a dynamic web framework.
Iraq War:
Full EMR: Working with the Veterans Affairs and Dept. of Defense, the ESSENCE team investigates the benefits of and issues with including elements of an EMR beyond the demographics and chief complaints. New algorithms that detect the severity of patients to better characterize an outbreak are built. Working with multi-terabyte database sizes prepares the ESSENCE system for dealing with large datasets in every jurisdiction.
After the Boston bombings, ESSENCE was adapted to look for non-infectious health events including: Anxiety, Depression, Suicidal Tendencies, and Hearing Loss. The system helped track the health of the Boston communities as they dealt with the aftermath of the attacks.
The Cloud version of ESSENCE
ESSENCE could not have been built without the support of many sponsors, numerous collaborators, perceptive users, and all of the dedicated members of the ESSENCE team.
APL

The Evolution of ESSENCE

Richard Wojcik, Howard Burkom, Joseph Lombardo

The Centers for Disease Control and Prevention announced that the National Syndromic Surveillance Program (NSSP) is transitioning to new syndromic surveillance software—the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE)—which will be hosted on the cloud-based BioSense Platform.

The Centers for Disease Control and Prevention announced that the National Syndromic Surveillance Program (NSSP) is transitioning to new syndromic surveillance software—the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE)—which will be hosted on the cloud-based BioSense Platform.

After the Boston bombings, ESSENCE was adapted to look for non-infectious health events including: Anxiety, Depression, Suicidal Tendencies, and Hearing Loss. The system helped track the health of the Boston communities as they dealt with the aftermath of the attacks.

The Cloud version of ESSENCE

ESSENCE could not have been built without the support of many sponsors, numerous collaborators, perceptive users, and all of the dedicated members of the ESSENCE team.

Current ESSENCE Locations

State
- Washington
- Oregon
- Indiana
- Nebraska
- Texas
- Maryland
- District of Columbia
- Virginia
- Florida
- Arkansas
- Tennessee
- Delaware

ESSENCE Locations

Regional, County, City
- Missouri & St. Louis, IL
- Aggregate NCR

- Tri-County, CO
- Stanislaus County, CA
- Santa Clara County, CA
- Cook County, IL
- Tarrant County, TX
- Marion County, IN

- Oklahoma City, OK
- Boston, MA

Veterans Affairs

DoD

National Syndromic Surveillance Program

Jan 2016

Acknowledgement

- We gratefully acknowledge the CDC NSSP team for their support of ESSENCE and this training effort

Basic ESSENCE System Components

Hands-on Guide

Content was developed for and funded by the Centers for Disease Control and Prevention (CDC) for training purposes. The findings and conclusions in this presentation are those of the authors and do not necessarily represent the views of CDC.

Center for Surveillance, Epidemiology, and Laboratory Services
Division of Health Informatics and Surveillance

ESSENCE Training Workshop

Basic System Components

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

Introduction

This is the first of two stepwise laboratory exercises that guides the user through select ESSENCE features and functions. Initially, it is recommended that users follow the suggested paths to walk through the basic components of ESSENCE. However, soon it will become evident that there is more than one pathway to access ESSENCE data visualization and analysis features.

Given that there is no one single “correct” method for using ESSENCE, after walking through suggested paths within this exercise, the user is encouraged to further explore additional functions embedded within ESSENCE features. With frequent use and familiarity, over time, individuals often establish their preferred path(s) for viewing ESSENCE visualizations and analysis outputs of interest.

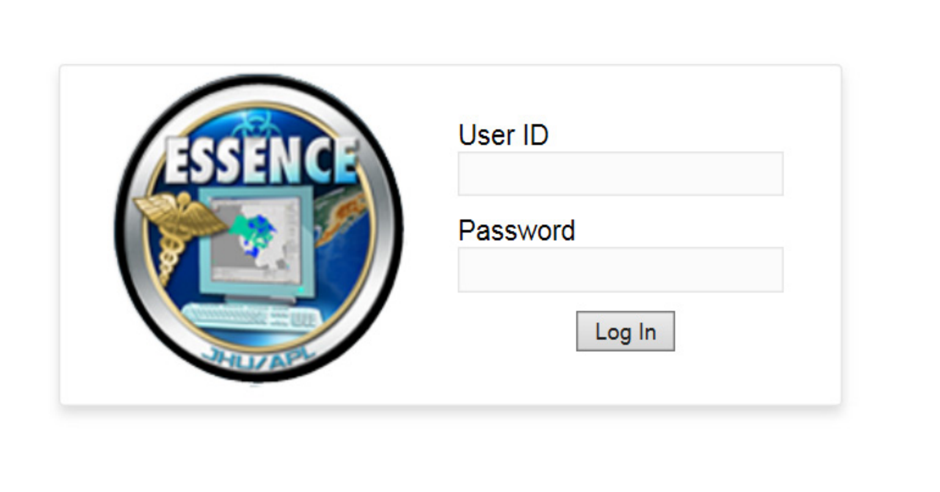
Features and Functions

Within this challenge, you will:

- **Log into ESSENCE**
- **Access the Query Portal and conduct a simple query**
  - **Use the following functions on the Time Series Page :**
    - **Weekly Time Series Viewer**
    - **Stacked Graphs**
    - **Detector Comparisons**
    - **Configuration Options**
    - **Stratification Queries**
    - **Overlay**
  - **View a Data Details page**
  - **View a GIS Map**
- **Access the Alerts Lists and get familiar with the options and fields**

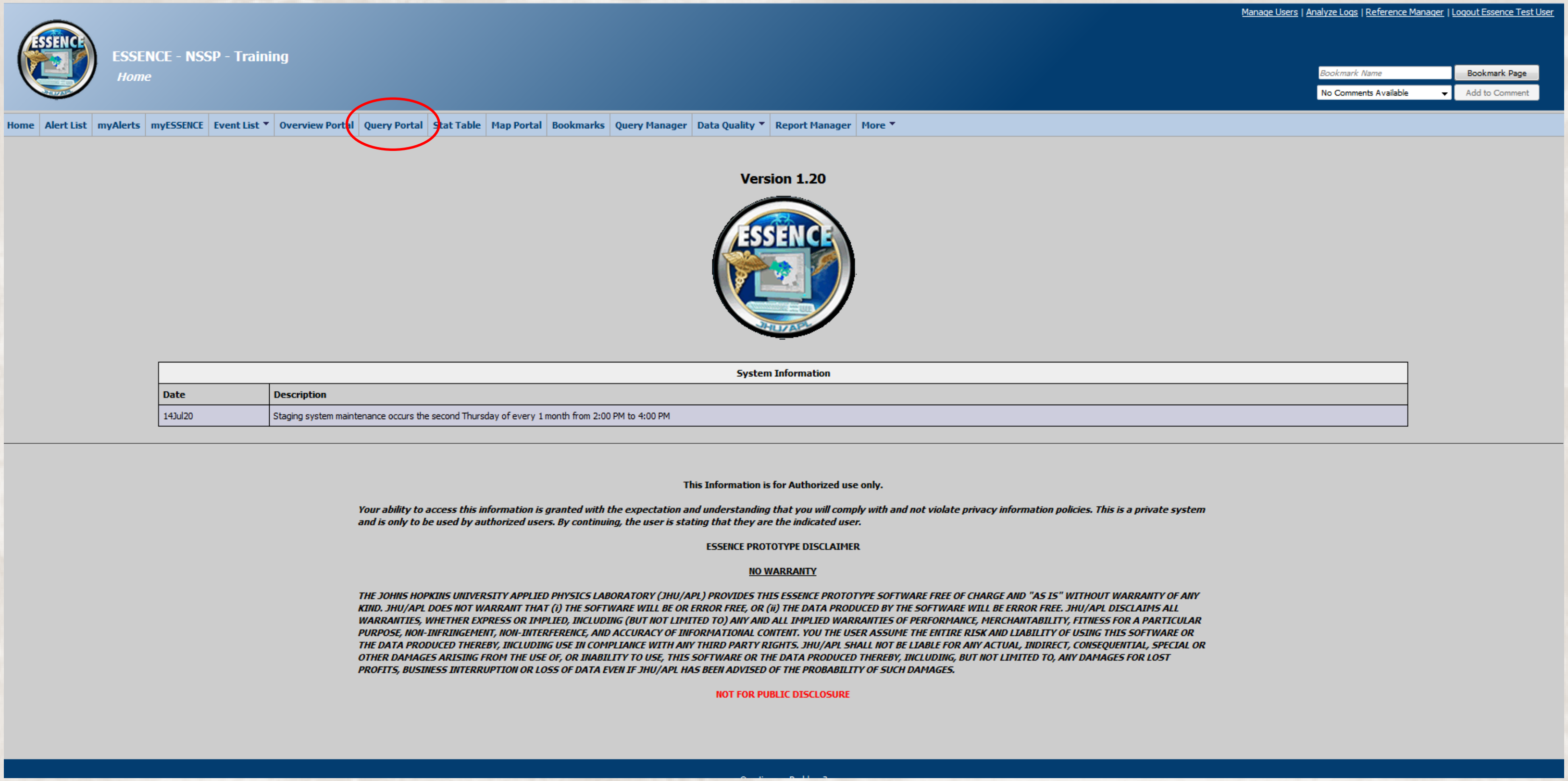
Logging In

- **Log into the NSSP ESSENCE training site:**
  - **Go to https://cloudessence.jhuapl.edu/nssp_essence**
- **Note:** ***Mozilla Firefox is the recommended web browser for use with ESSENCE. Compatibility is not guaranteed with other browsers.***
  - **Enter your user ID and password and click the Log In button**



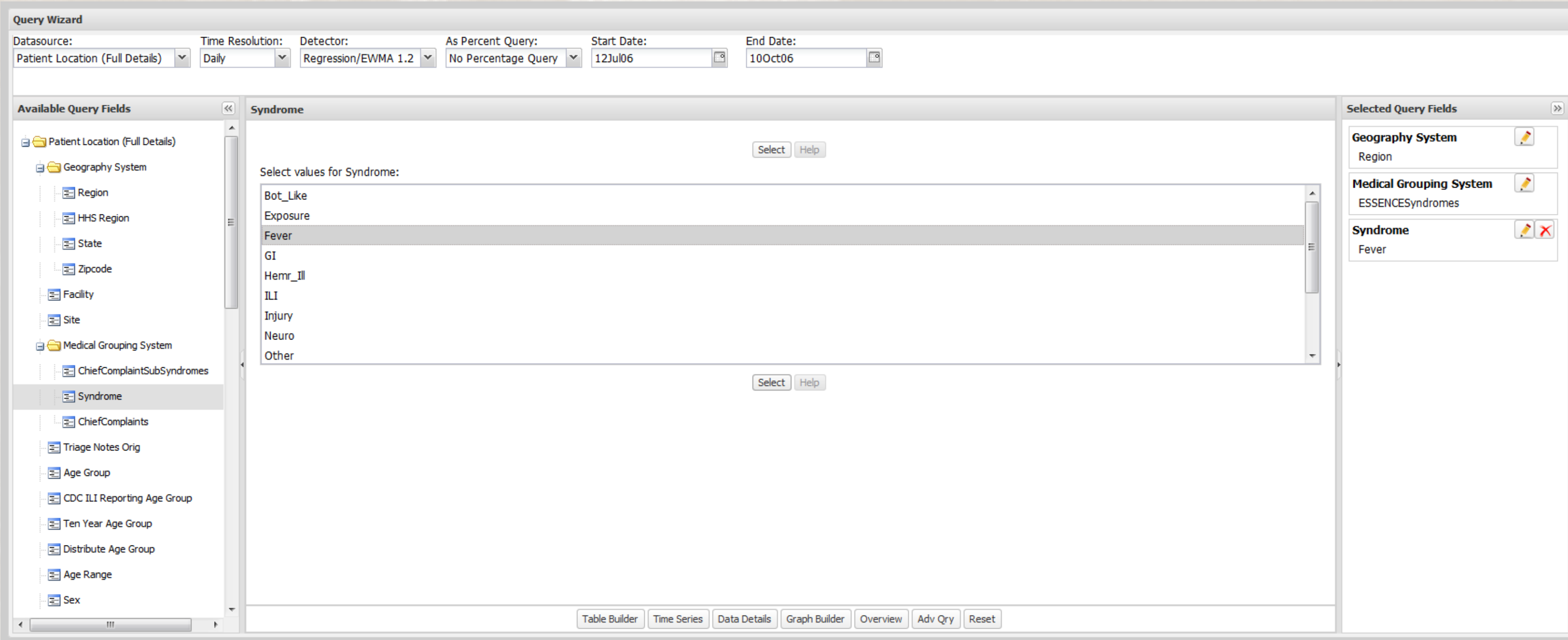
  - **This will take you to the NSSP ESSENCE home page**

Accessing the Query Portal



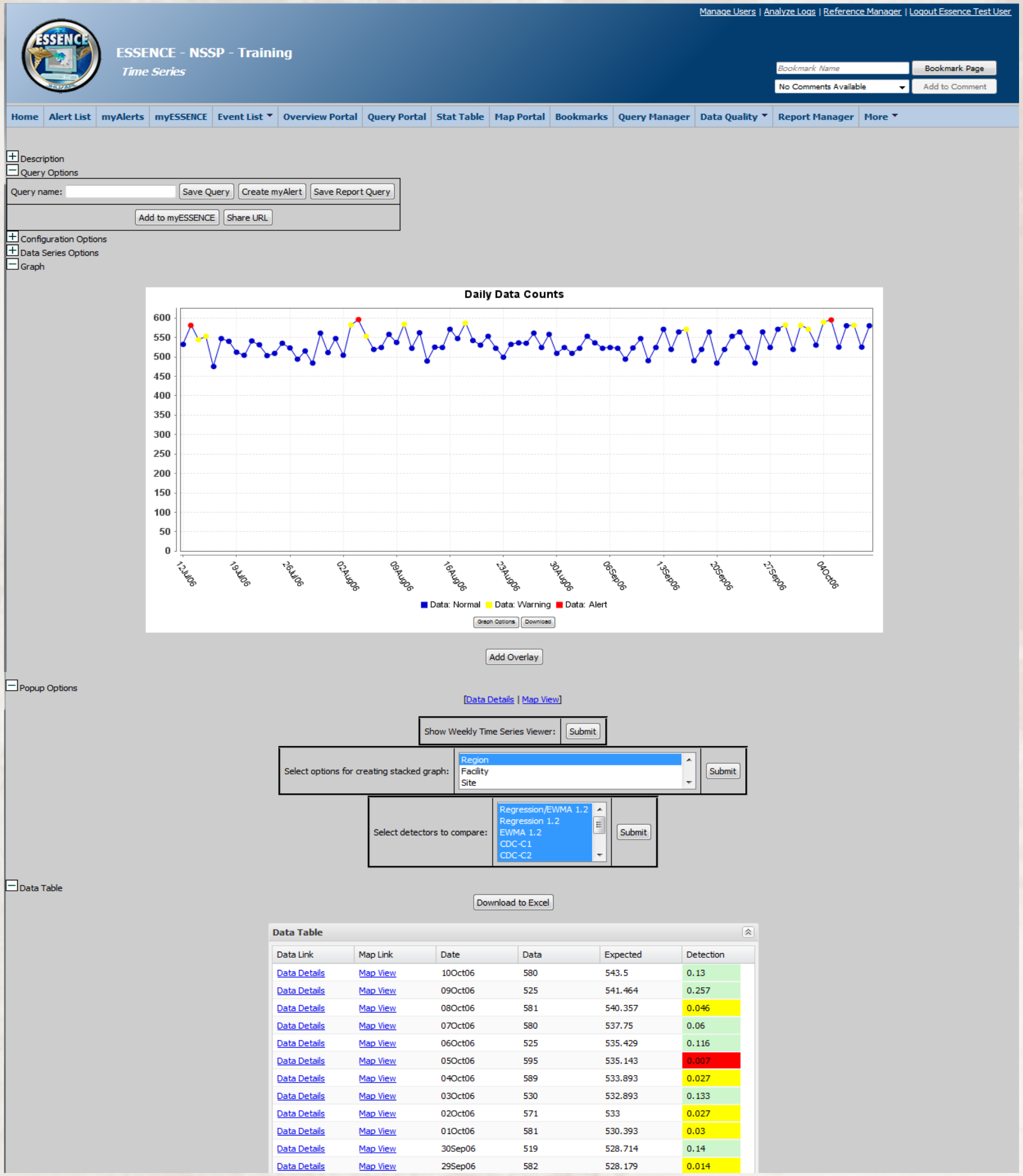
- **The Home Page provides access to the System Information section. This section can contain announcements and information posted by the administrators.**
- **For this walkthrough, please click on the Query Portal**

Using the Query Portal



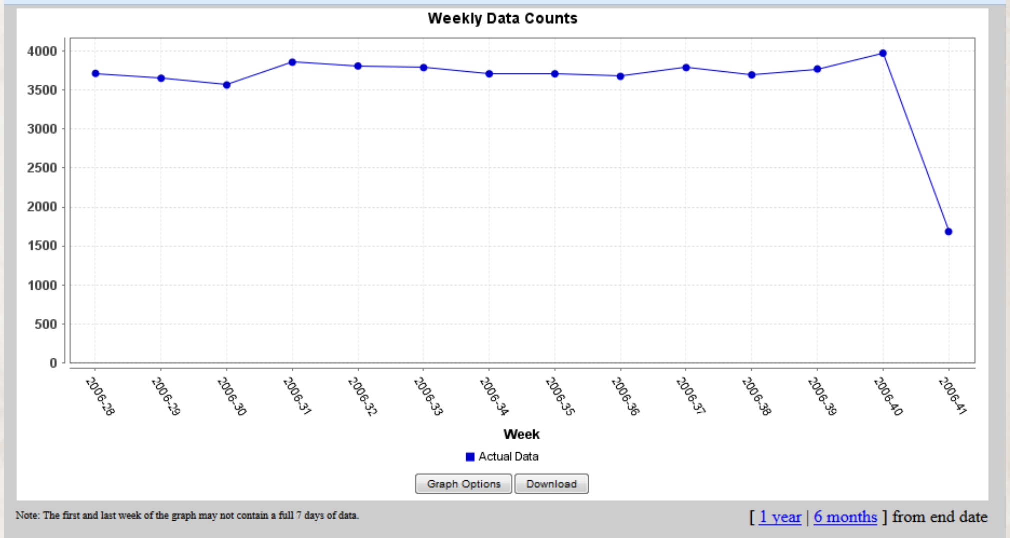
- **To use the Query Portal, you will first choose your Datasource.**
- **Then choose any Time Resolution, Detector, Date Range, and if you want a percent query option**
- **Next you can select any parameter on the left pane. It will fill in the center pane with the available options for you to choose for that parameter.**
- **Once you have chosen all your parameters, choose the ESSENCE feature you want to use your query definition in: Table Builder, Time Series, Data Details, Graph Builder, Overview**
- **If you need to create a more complex query using and/or logic between parameters, you can choose the Advanced Query Tool option from this bottom at any time.**
- **For this walkthrough, we will choose the Time Series Option next.**

Time Series Page



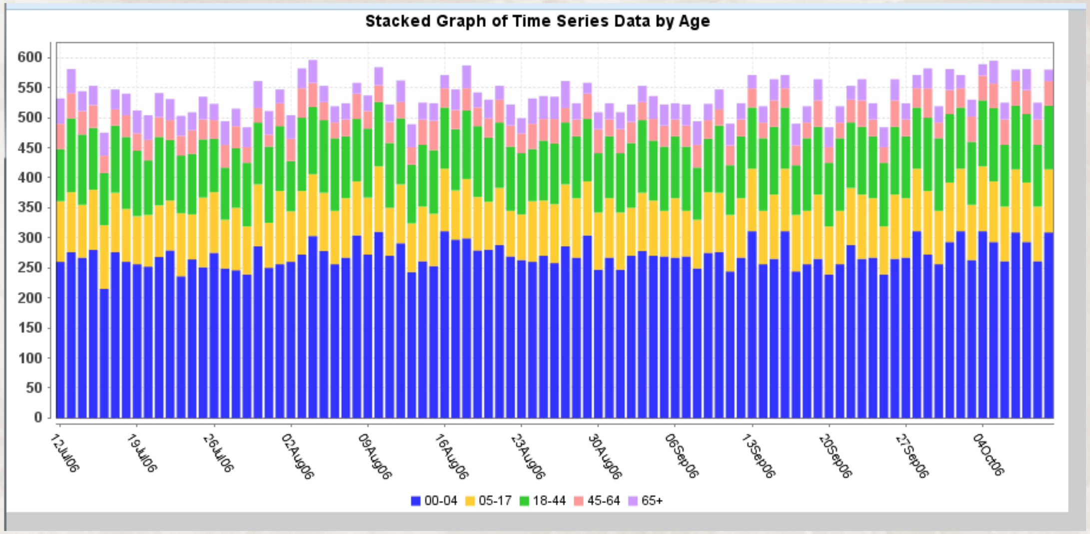
- You can view your time series image and mouse over any point to get more information.
- You can view the data from the query in the Data Table including the count, expected value from the detector, and detector output (normally Pvalue)
- You can view popup graphs showing stacked graphs, weekly views, and detector comparison plots.
- You can perform an overview query and apply it directly to your existing graph.
- You can save this query / time series for use in myAlerts, myESSENCE, or your saved Query Manager.
- You can stratify your query under the Data Series Options to view a breakdown of parameters, such as Age Group or Geographic Region.
- For the purpose of this walkthrough, please click on the “Show Weekly Time Series Viewer” submit button.

Time Series Page: Weekly Popup



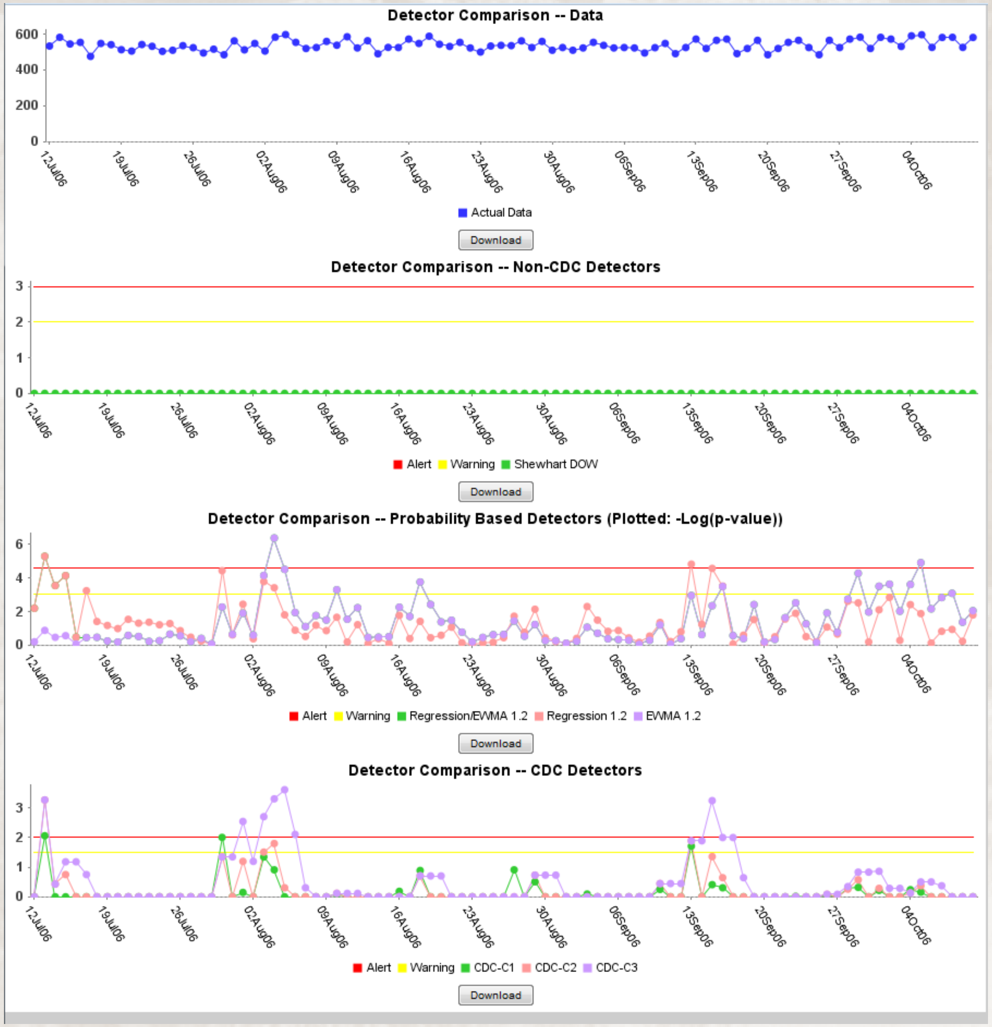
- This popup will show the query in a weekly form.
- You can modify the date range quickly by choosing 1 year or 6 months options underneath the graph.
- For the purpose of this walkthrough, please close this window and choose an “Age Group” stacked graph popup option.

Time Series Page: Weekly Popup



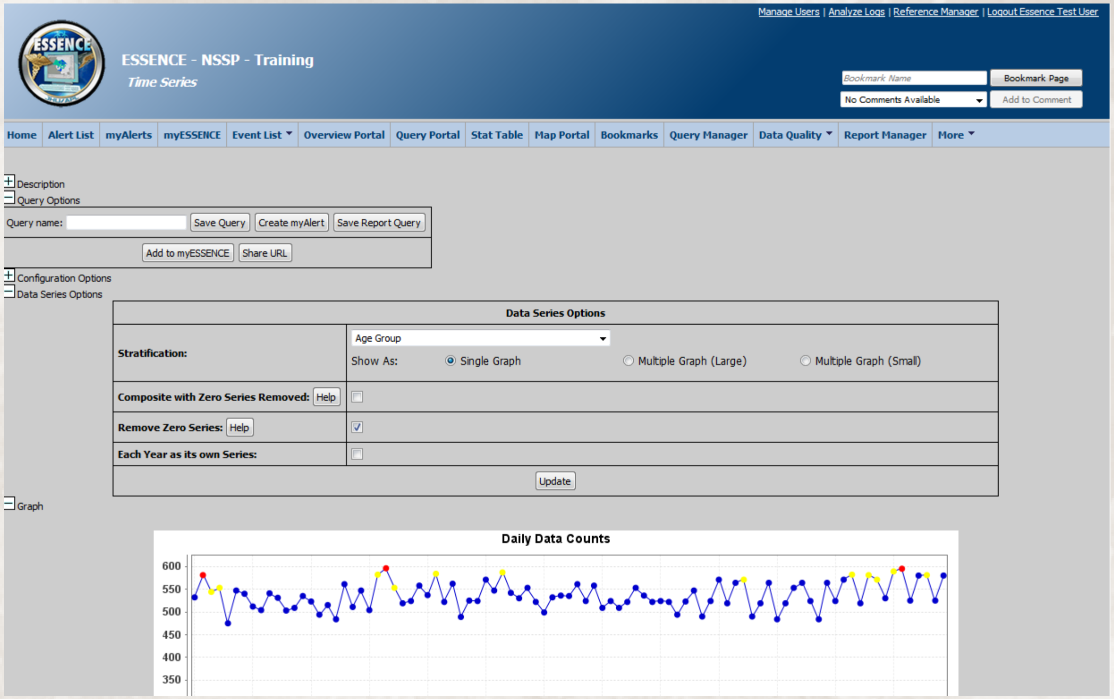
- **This popup will show the query broken down by the parameter chosen in a stacked graph.**
- **You can mouse over the graph to get additional details.**
- **For the purpose of this walkthrough, please close this window and choose the Submit button on the "Select detectors to compare" popup.**

Time Series Page: Weekly Popup



- This popup will show the query in the top graph, Non-CDC algorithms in the middle graph, and the CDC Ears algorithms in the bottom graph.
- This allows users to compare the results of multiple detectors at one time.
- For the purpose of this walkthrough, please close this window and follow the instructions on the next slide.

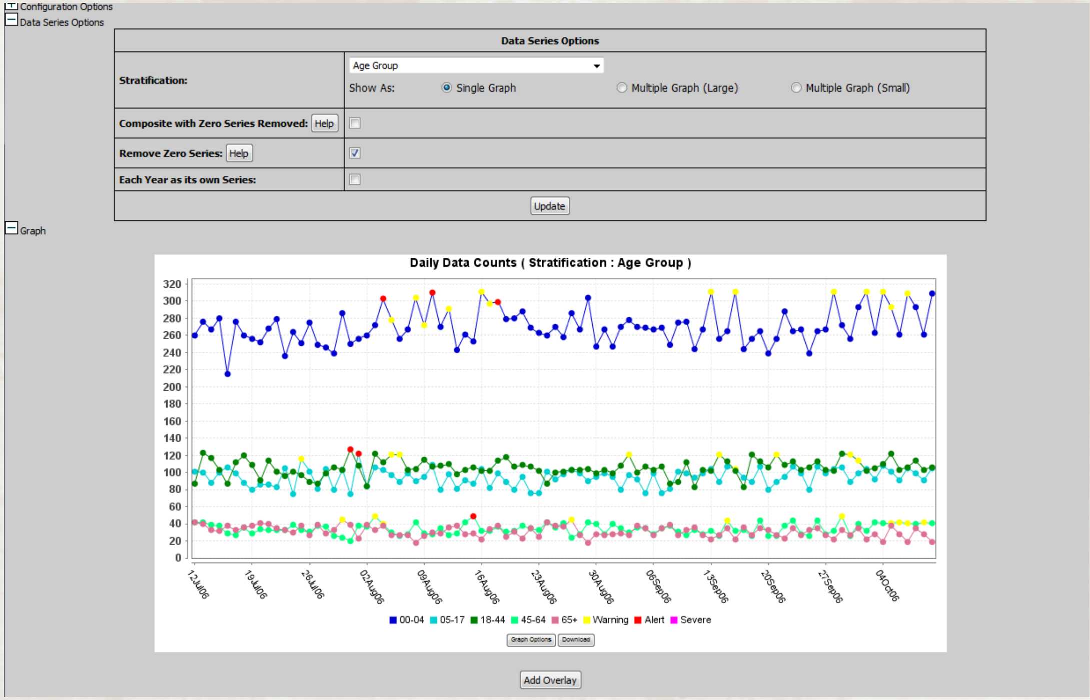
Time Series Page: Data Series Options



The composite feature runs detection on the sum of the data from each series based on a predefined stratification (e.g. Hospital, SchoolID, StoreID). It removes any series from the sum that contains one or more zero values. This includes any zero in the entire baseline plus the additional time prior to the start date used to warm-up the detectors (around 40 days).

- **Under the Data Series Options, you can choose a parameter to stratify by.**
- **These stratification queries can be shown in a single graph (if number of series is small enough), multiple graphs (large) and multiple graphs (small).**
- **There are also options for Composite Detection, Removing Zero Series, and putting each year as its own series.**
- **For the purpose of this walkthrough, please click Age Group, and Update.**

Time Series Page: Stratification Graph

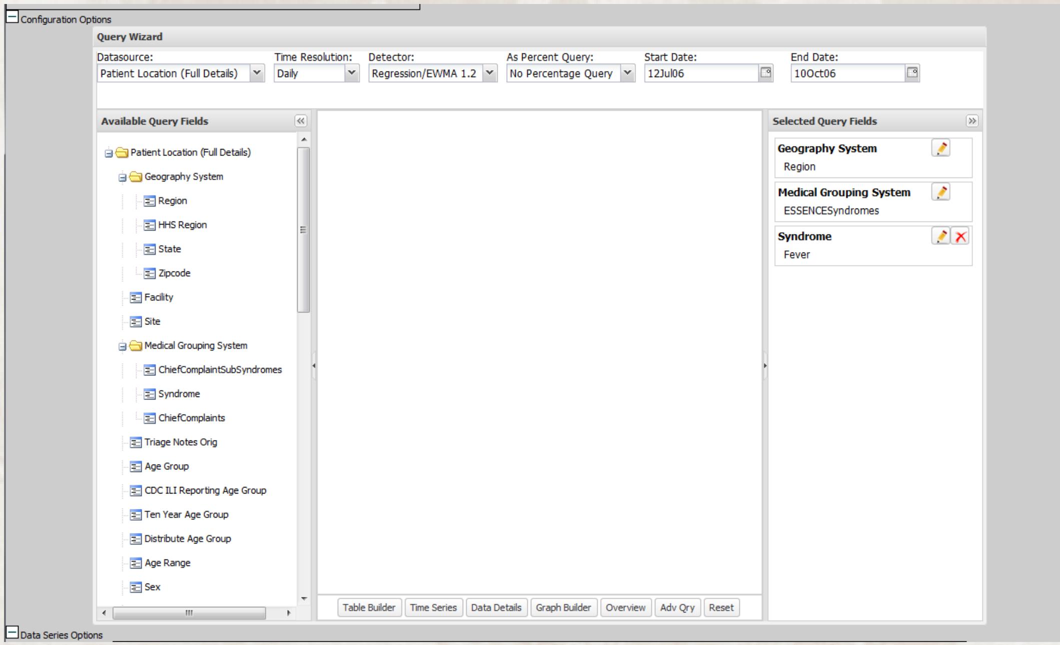


- The Stratification Graph will contain detector results from each series.
- For the purpose of this walkthrough, please click on the Show As option: Multiple Graph (Small) and click Update.

Time Series Page: Stratification Graph

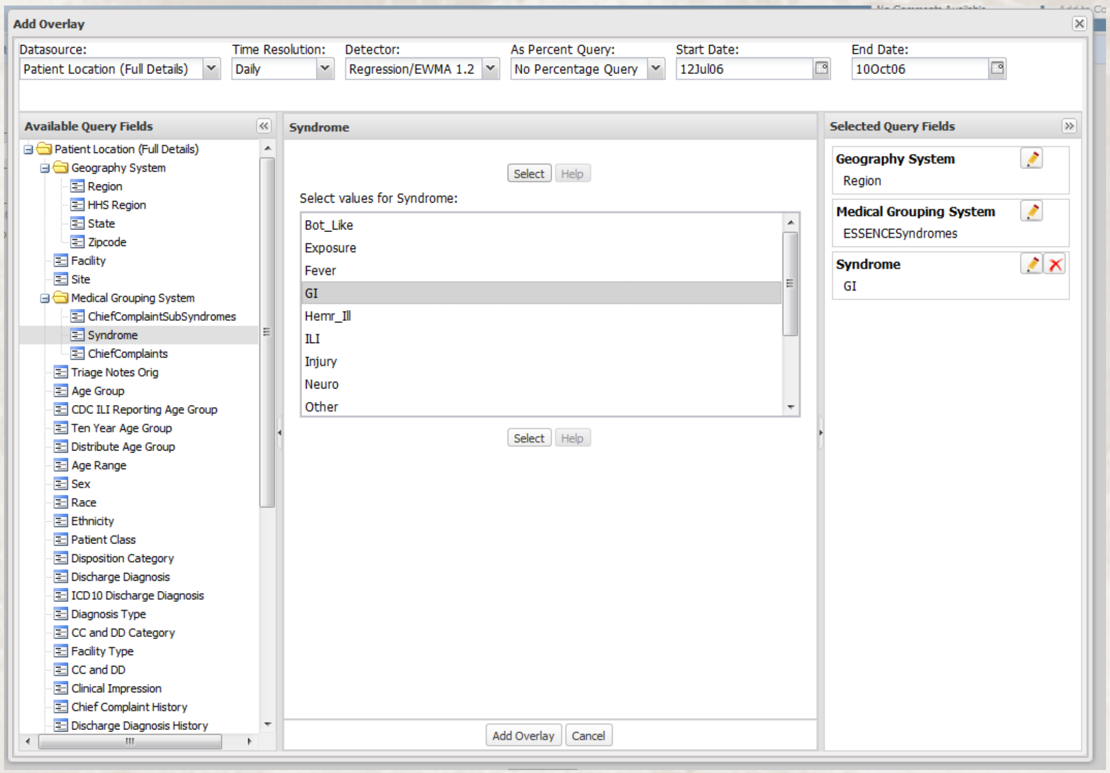
- Each series is now in its own graph.
- For the purpose of this walkthrough, please click on the plus sign next to the Configuration Option label.

Time Series Page: Configuration Options



- **From many locations in ESSENCE, you can change the definition of the query you are currently looking at by choosing the Configuration Options.**
- **Additionally, on the time series page, you can undo all the stratifications or overlays you have performed by clicking on the Time Series button again.**
- **For the purpose of this walkthrough, please click on the Time Series button, then click on the Overlay button.**

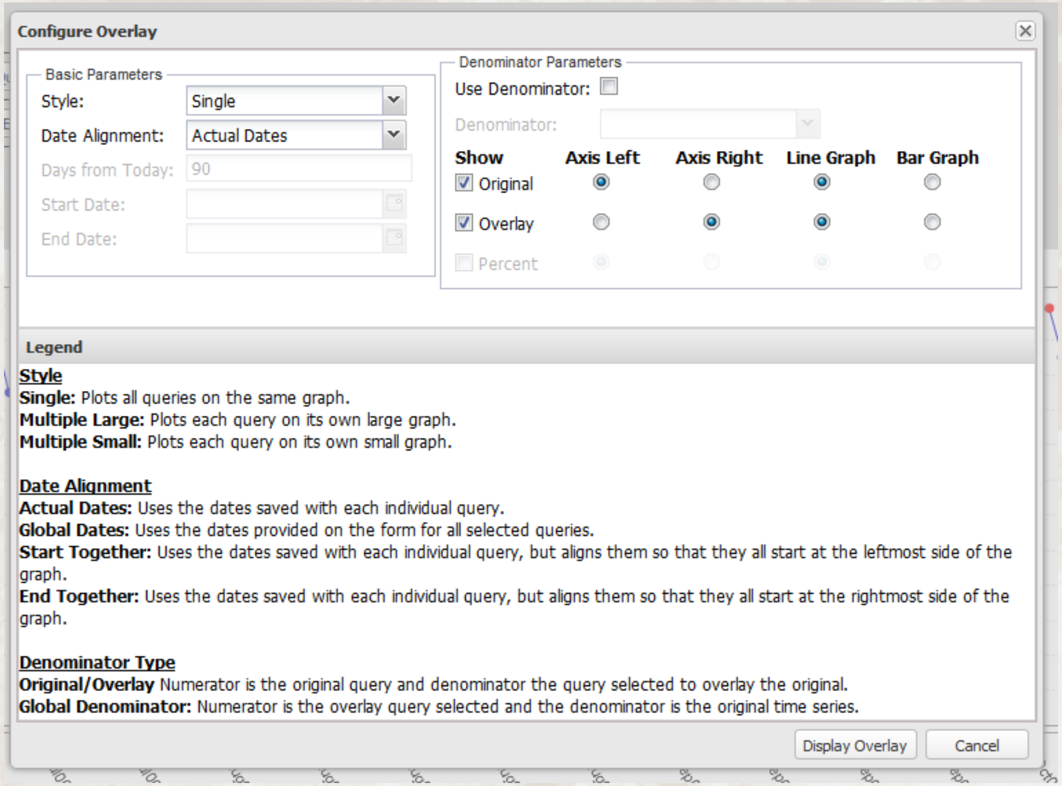
Time Series Page: Overlay



- The Overlay option will allow you to create a new query, and overlay it on top of the existing "original" query you performed.

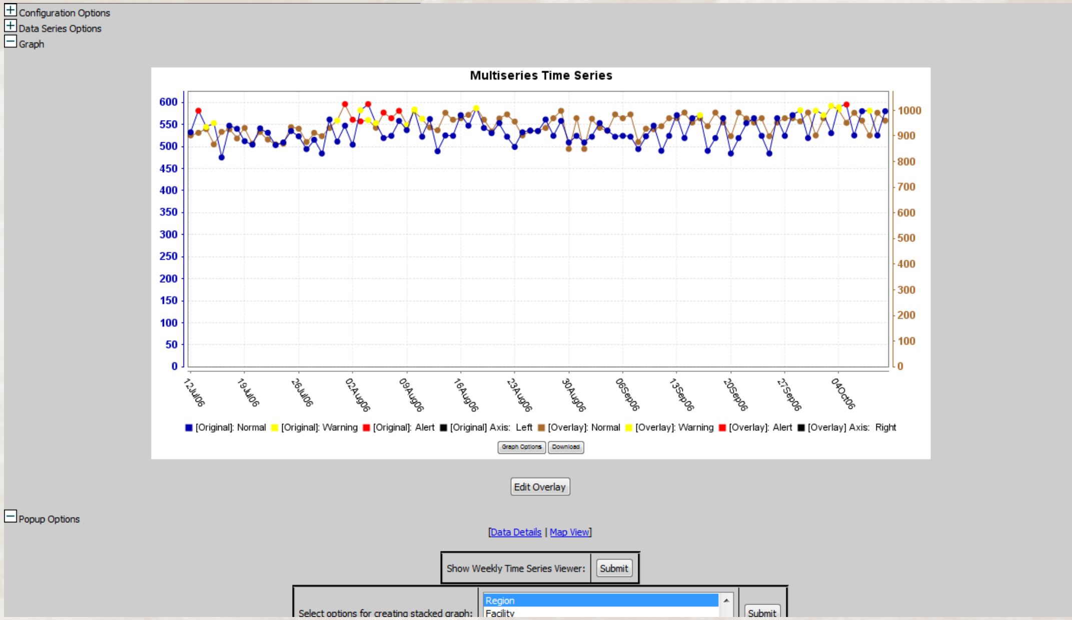
- For the purpose of this walkthrough, define a new query that is different from your original query, and click "Add Overlay".

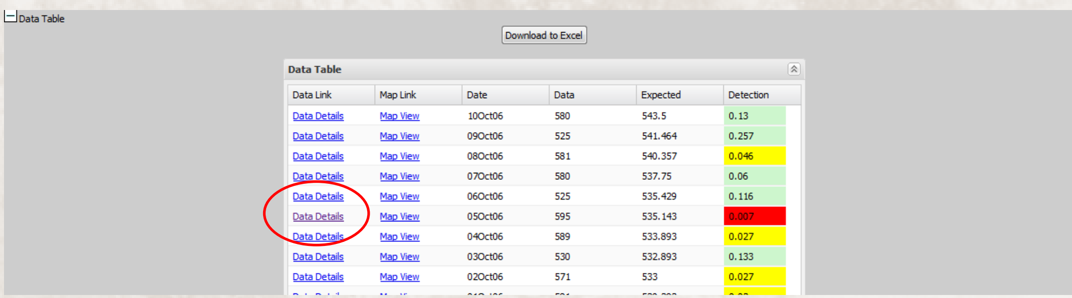
Time Series Page: Overlay



- **The Overlay configuration window, you can choose single or multiple graphs, and date alignment.**
- **Under the denominator parameters section, you can decide if you want to have the one of the queries divided by the other.**
- **You can also display the overlay and/or original query on the same or different axis.**
- **For the purpose of this walkthrough, leave the defaults and click Display Overlay.**

Time Series Page: Overlay





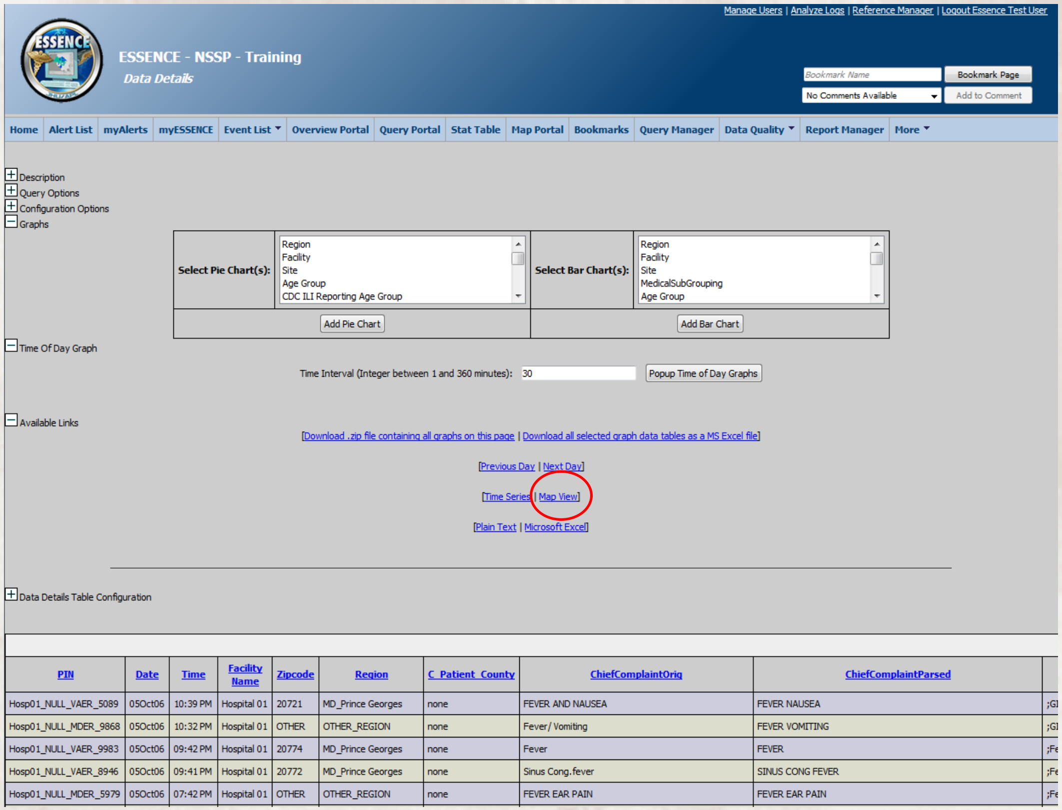
Data Table

Data Link	Map Link	Date	Data	Expected	Detection
Data Details	Map View	10Oct06	580	543.5	0.13
Data Details	Map View	09Oct06	525	541.464	0.257
Data Details	Map View	08Oct06	581	540.357	0.046
Data Details	Map View	07Oct06	580	537.75	0.06
Data Details	Map View	06Oct06	525	535.429	0.116
Data Details	Map View	05Oct06	595	535.143	0.007
Data Details	Map View	04Oct06	589	533.893	0.027
Data Details	Map View	03Oct06	530	532.893	0.133
Data Details	Map View	02Oct06	571	533	0.027

- The result will be displayed.
- Currently, the data table below the graph only represents the original query. We hope to update this in the future to include both the original and the overlay.

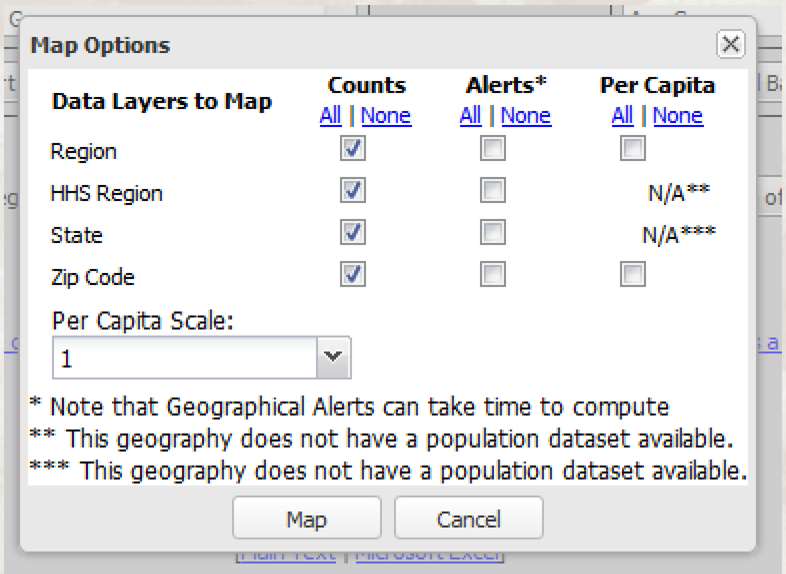
- For the purpose of this walkthrough, click on a data details link in the data table below for a single date.

Data Details Page



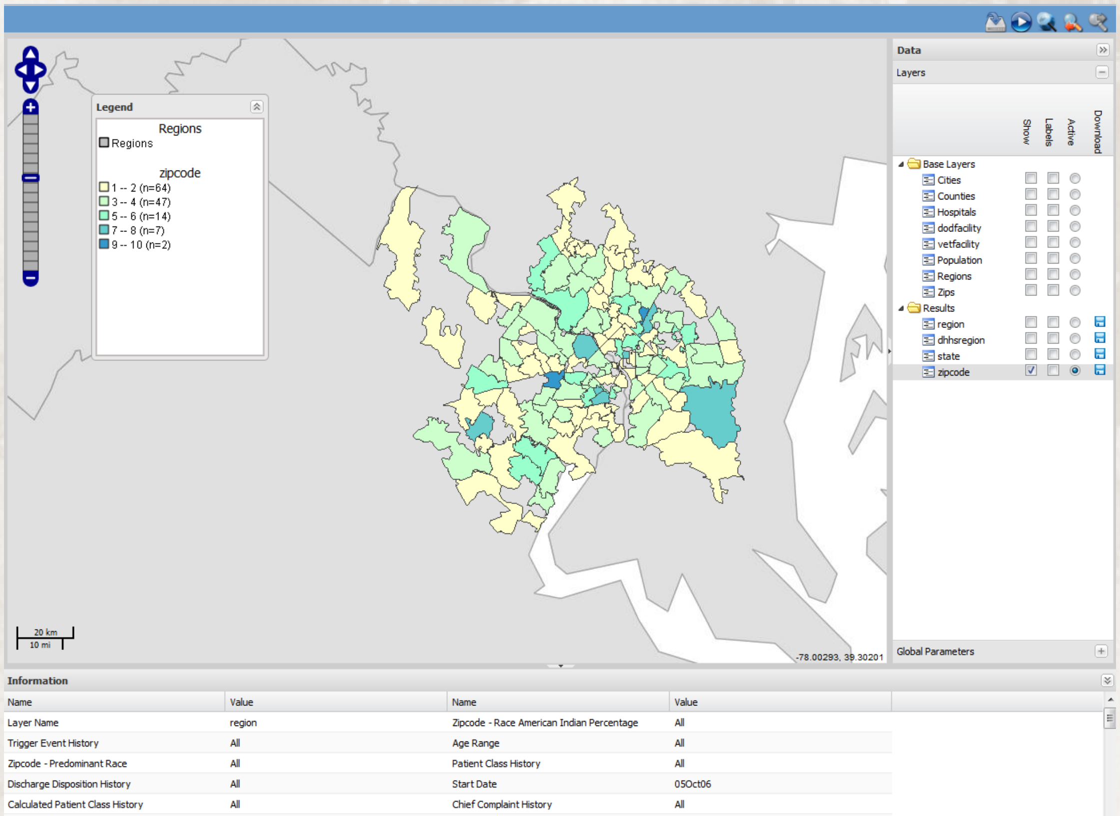
- The Data Details provides the line listings for the query you performed.
- You can scroll left/right to view all the information provided by that data source.
- You can select Pie/Bar charts to view breakdowns of individual parameters.
- You can download the information in CSV or Excel formats.
- You can view the information broken down by 30/60/90/120 minute windows.
- You can control which columns are visible to your account in the Data Details Table Configuration.
- You can sort by clicking on a column header.
- For this walkthrough, please click on the Map View link.

Map View



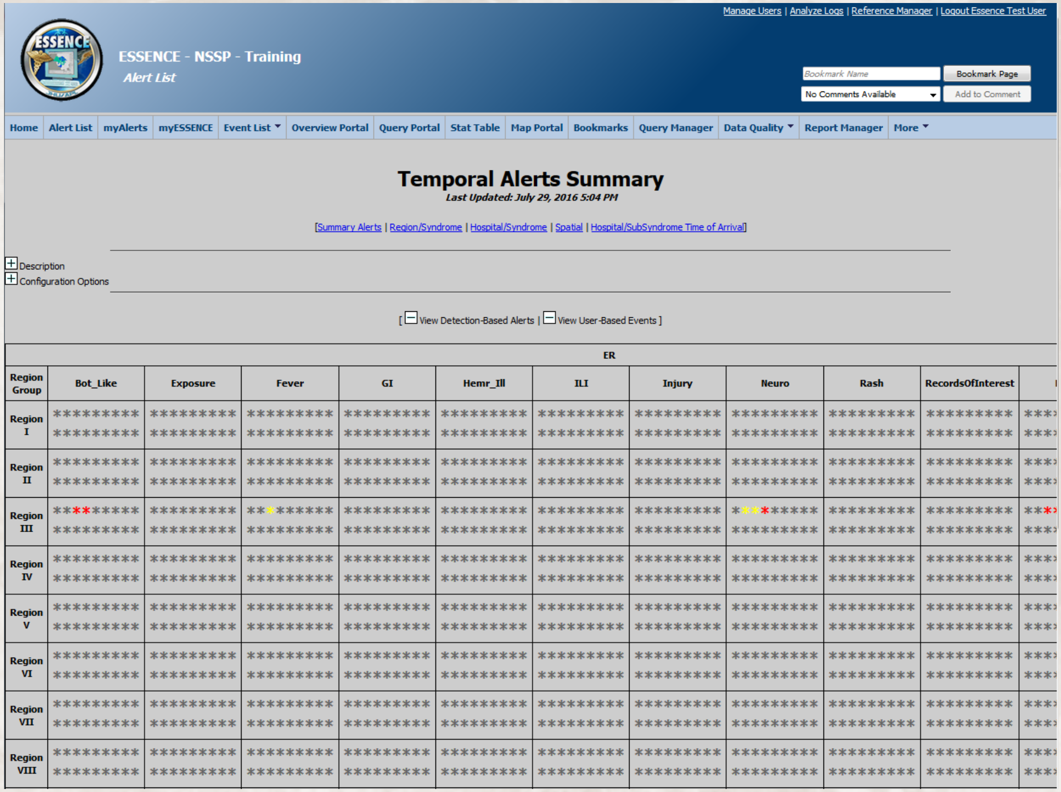
- When you click on a Map View link, you are given these options.
- For this walkthrough, leave the default options checked, and click Map.

Map View



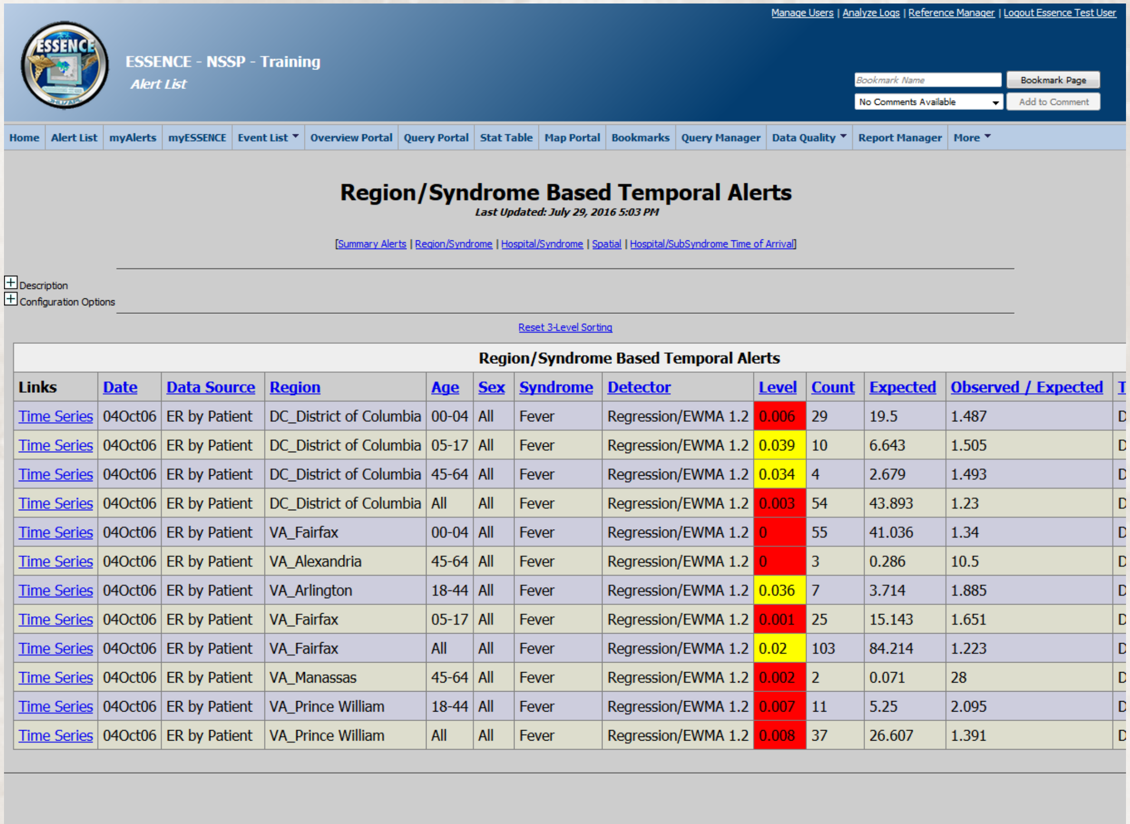
- The Map View allows you to zoom / pan to see any part of the map.
- You can make layers visible / invisible by checking the “Show” box next to a layer’s name.
- You can make labels visible / invisible by checking the “Labels” box next to a layer’s name.
- The active layer is the layer that will be selected if using any selection tools.
- There are tools in the upper right corner that allow you to save a Map to be used in a report (and make it easier to download the image or print). There is also a tool to allow you to create an animated movie of the map over time.
- The bottom of the map will show you information about the query or what is currently selected.
- Special note: If you cannot see your layer, it may be hidden underneath another already visible layer. Click the active button to bring it to the top.
- For this walkthrough, please close the Map window and click on the Alert List menu option.

Alert List: Summary



- The Summary Alert List is made up of 2 rows of stars in each Region Group / Syndrome cell.
- The stars represent the last 9 days (most recent day to the right), and are color coded.
- The top row represents mathematical alerts from the Region / Syndrome Temporal Alerts page.
- The bottom row represents concern levels discussed by users in the Event List.
- Note: A grey cell does not mean there are zero Region / Syndrome Alerts. It just means that there were either not enough or none strong enough to create a Summary Level alert.
- For this walkthrough, please click on a Fever Summary Alert.

Alert List: Region / Syndrome Temporal Alerts


Manage Users | Analyze Logs | Reference Manager | Logout Essence Test User

ESSENCE - NSSP - Training

Alert List

Bookmark Page

No Comments Available

Add to Comment

Home | Alert List | myAlerts | myESSENCE | Event List | Overview Portal | Query Portal | Stat Table | Map Portal | Bookmarks | Query Manager | Data Quality | Report Manager | More

Region/Syndrome Based Temporal Alerts

Last Updated: July 29, 2016 5:03 PM

[Summary Alerts | Region/Syndrome | Hospital/Syndrome | Spatial | Hospital/SubSyndrome Time of Arrival]

Description

Configuration Options

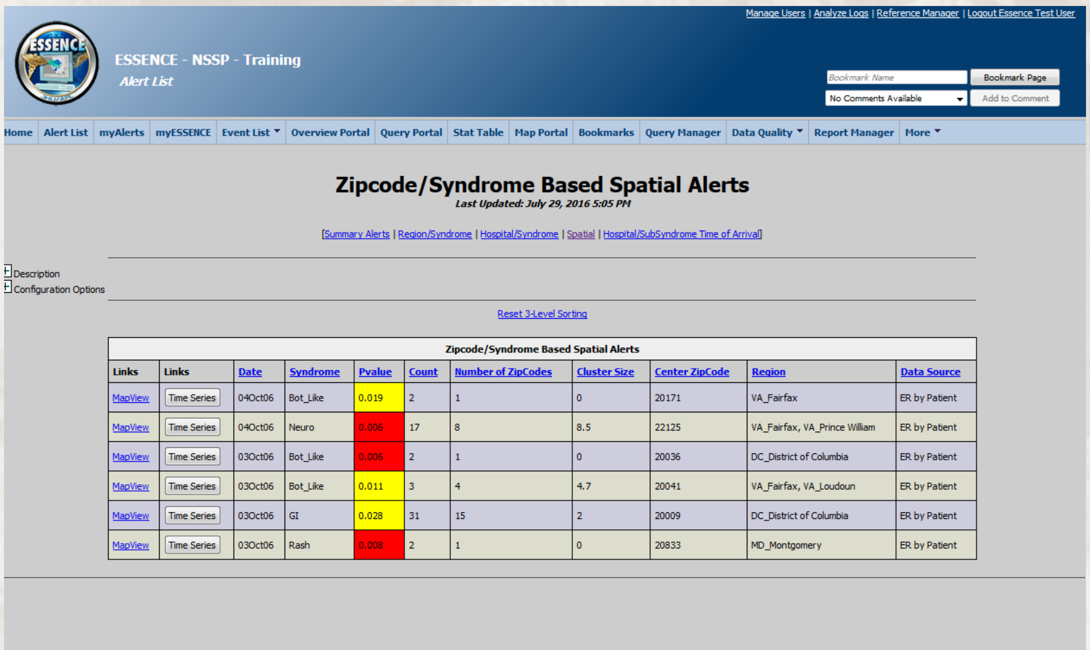
Reset 3-Level Sorting

Region/Syndrome Based Temporal Alerts

Links	Date	Data Source	Region	Age	Sex	Syndrome	Detector	Level	Count	Expected	Observed / Expected	T
Time Series	04Oct06	ER by Patient	DC_District of Columbia	00-04	All	Fever	Regression/EWMA 1.2	0.006	29	19.5	1.487	D
Time Series	04Oct06	ER by Patient	DC_District of Columbia	05-17	All	Fever	Regression/EWMA 1.2	0.039	10	6.643	1.505	D
Time Series	04Oct06	ER by Patient	DC_District of Columbia	45-64	All	Fever	Regression/EWMA 1.2	0.034	4	2.679	1.493	D
Time Series	04Oct06	ER by Patient	DC_District of Columbia	All	All	Fever	Regression/EWMA 1.2	0.003	54	43.893	1.23	D
Time Series	04Oct06	ER by Patient	VA_Fairfax	00-04	All	Fever	Regression/EWMA 1.2	0	55	41.036	1.34	D
Time Series	04Oct06	ER by Patient	VA_Alexandria	45-64	All	Fever	Regression/EWMA 1.2	0	3	0.286	10.5	D
Time Series	04Oct06	ER by Patient	VA_Arlington	18-44	All	Fever	Regression/EWMA 1.2	0.036	7	3.714	1.885	D
Time Series	04Oct06	ER by Patient	VA_Fairfax	05-17	All	Fever	Regression/EWMA 1.2	0.001	25	15.143	1.651	D
Time Series	04Oct06	ER by Patient	VA_Fairfax	All	All	Fever	Regression/EWMA 1.2	0.02	103	84.214	1.223	D
Time Series	04Oct06	ER by Patient	VA_Manassas	45-64	All	Fever	Regression/EWMA 1.2	0.002	2	0.071	28	D
Time Series	04Oct06	ER by Patient	VA_Prince William	18-44	All	Fever	Regression/EWMA 1.2	0.007	11	5.25	2.095	D
Time Series	04Oct06	ER by Patient	VA_Prince William	All	All	Fever	Regression/EWMA 1.2	0.008	37	26.607	1.391	D

- The Region / Syndrome alerts will provide a listing of all data slices (Datasource x Region x Age x Syndrome) that are alerting over the past 7 days (or on the day you chose from the Summary Alert List).
- For the default detector, the Level column contains the Pvalue.
- Each column can be sorted.
- Each alert can be investigated by clicking on the Time Series Link.
- For ease, it is common to right-click on the Time Series link and “Open in a new tab” to preserve your alert list window for further investigation.
- For this walkthrough, please click on the link for the Spatial alert list.

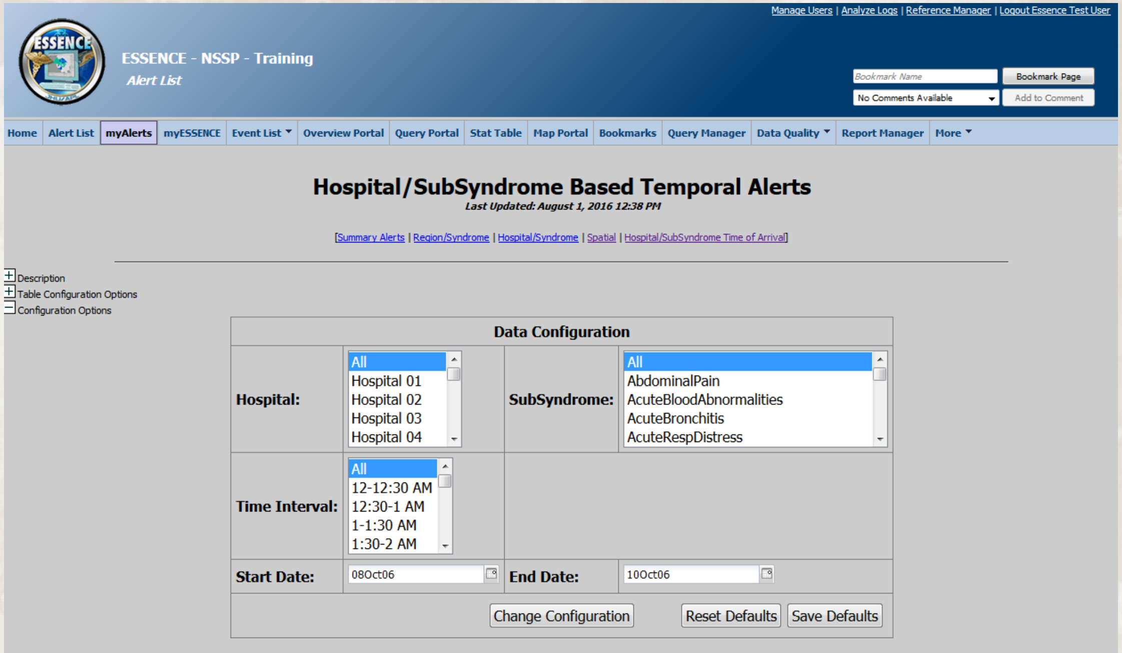
Alert List: Spatial

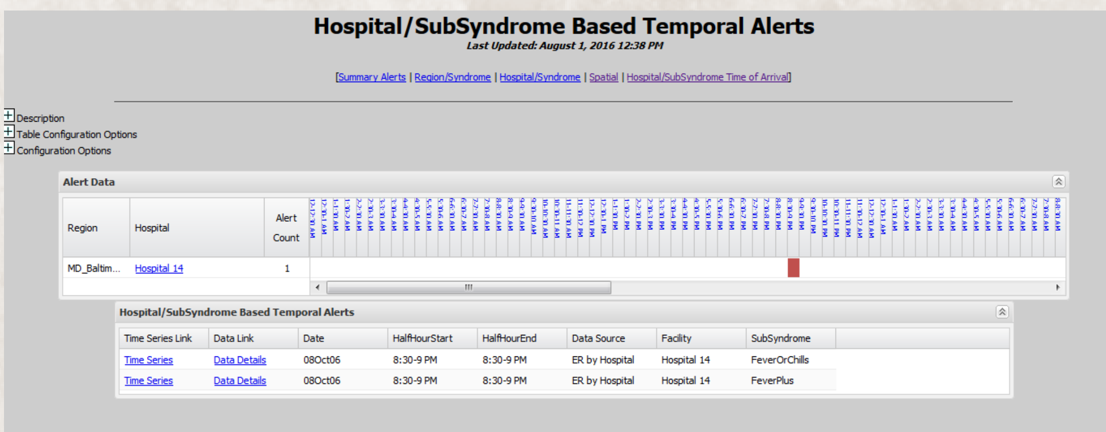


Links	Links	Date	Syndrome	Pvalue	Count	Number of ZipCodes	Cluster Size	Center ZipCode	Region	Data Source
MapView	Time Series	04Oct06	Bot_Like	0.019	2	1	0	20171	VA_Fairfax	ER by Patient
MapView	Time Series	04Oct06	Neuro	0.006	17	8	8.5	22125	VA_Fairfax, VA_Prince William	ER by Patient
MapView	Time Series	03Oct06	Bot_Like	0.006	2	1	0	20036	DC_District of Columbia	ER by Patient
MapView	Time Series	03Oct06	Bot_Like	0.011	3	4	4.7	20041	VA_Fairfax, VA_Loudoun	ER by Patient
MapView	Time Series	03Oct06	GI	0.028	31	15	2	20009	DC_District of Columbia	ER by Patient
MapView	Time Series	03Oct06	Rash	0.008	2	1	0	20833	MD_Montgomery	ER by Patient

- The Spatial Alert List will show any cluster alerts that have occurred in the past 8 days.
- The count is the number of cases.
- The cluster size is the diameter (in miles) of the zip code centroids involved in the cluster.
- The region is a comma separated list of the regions involved in the cluster.
- The Map View Link and Time Series button will allow you to investigate the cluster further.
- For this walkthrough, please click on the link for the Hospital / Subsyndrome Time of Arrival alert list.

Alert List: Time of Arrival (ToA)





- **To view ToA alerts, first choose your hospitals and subsyndromes of interest, then choose "Change Configuration"**
- **All ToA alerts will then be shown as red squares on the grid.**
- **If you click on any red square, a details table will be created to show all ToA alerts that fell into that Hospital / Time window.**
- **From there, you can click on Data Details or Time Series links that will allow you to investigate the alert further.**
- **This walkthrough is now complete.**

Advanced ESSENCE System Components

Hands-on Guide

Content was developed for and funded by the Centers for Disease Control and Prevention (CDC) for training purposes. The findings and conclusions in this presentation are those of the authors and do not necessarily represent the views of CDC.

Center for Surveillance, Epidemiology, and Laboratory Services
Division of Health Informatics and Surveillance

ESSENCE Training Workshop

Advanced System Components

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

Introduction

This is the second of two stepwise laboratory exercises that guides the user through select ESSENCE features and functions. Initially, it is recommended that users follow the suggested paths to walk through the basic components of ESSENCE. However, soon it will become evident that there is more than one pathway to access ESSENCE data visualization and analysis features.

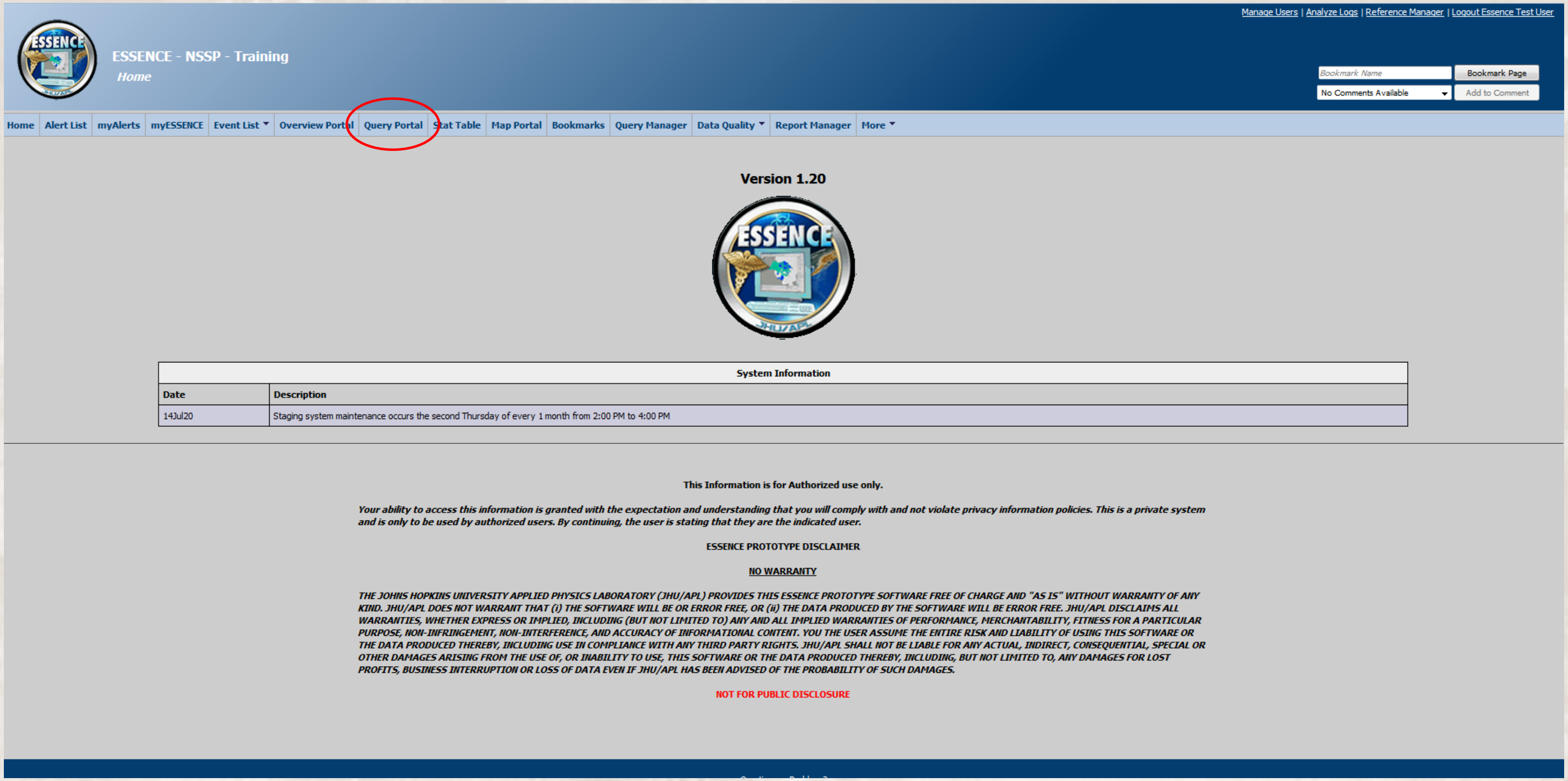
Given that there is no one single “correct” method for using ESSENCE, after walking through suggested paths within this exercise, the user is encouraged to further explore additional functions embedded within ESSENCE features. With frequent use and familiarity, over time, individuals often establish their preferred path(s) for viewing ESSENCE visualizations and analysis outputs of interest.

Features and Functions

Within this challenge, you will:

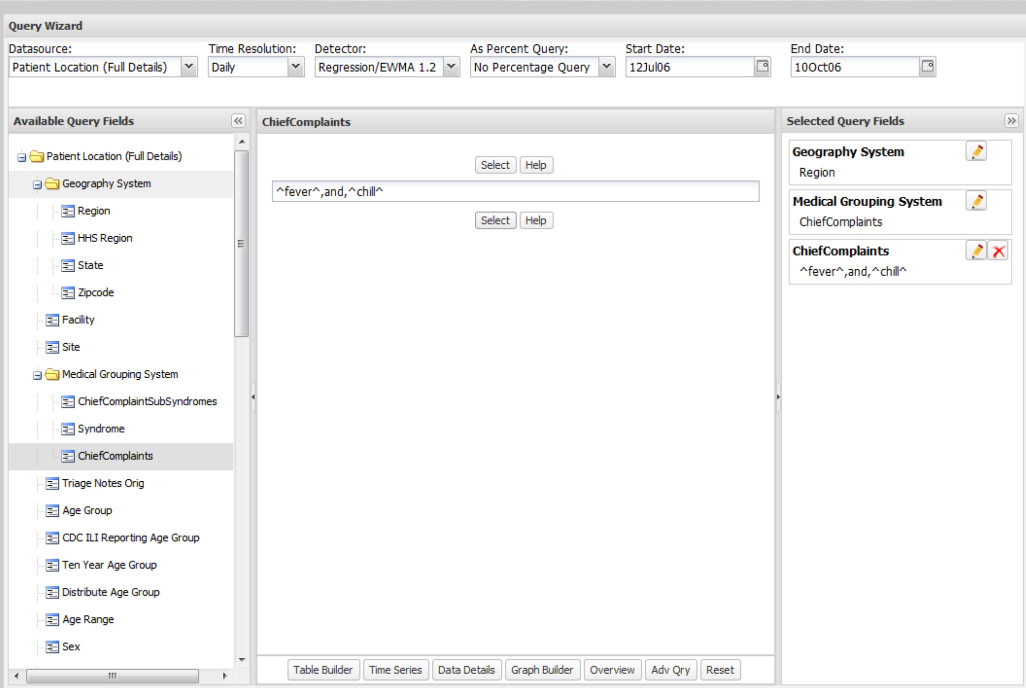
- Conduct a free-text query
- View advanced features of the Data Details Page
- Conduct an Advanced Query Tool (AQT) query
- Create and view myAlerts
- Create and view myESSENCE tabs
- Access Query Manager
- Access Report Manager
- Access the Overview Portal
- Access a Stat Table
- Access Data Quality Portal

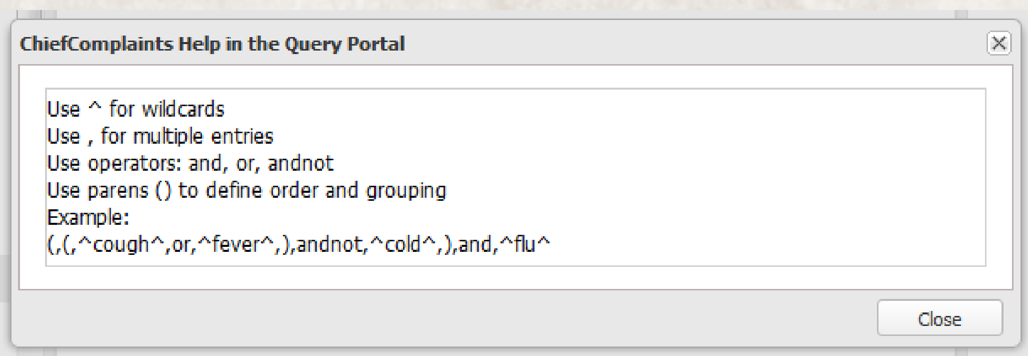
Accessing the Query Portal



- The Home Page provides access to the System Information section. This section can contain announcements and information posted by the administrators.
- For this walkthrough, please click on the Query Portal

Query Portal

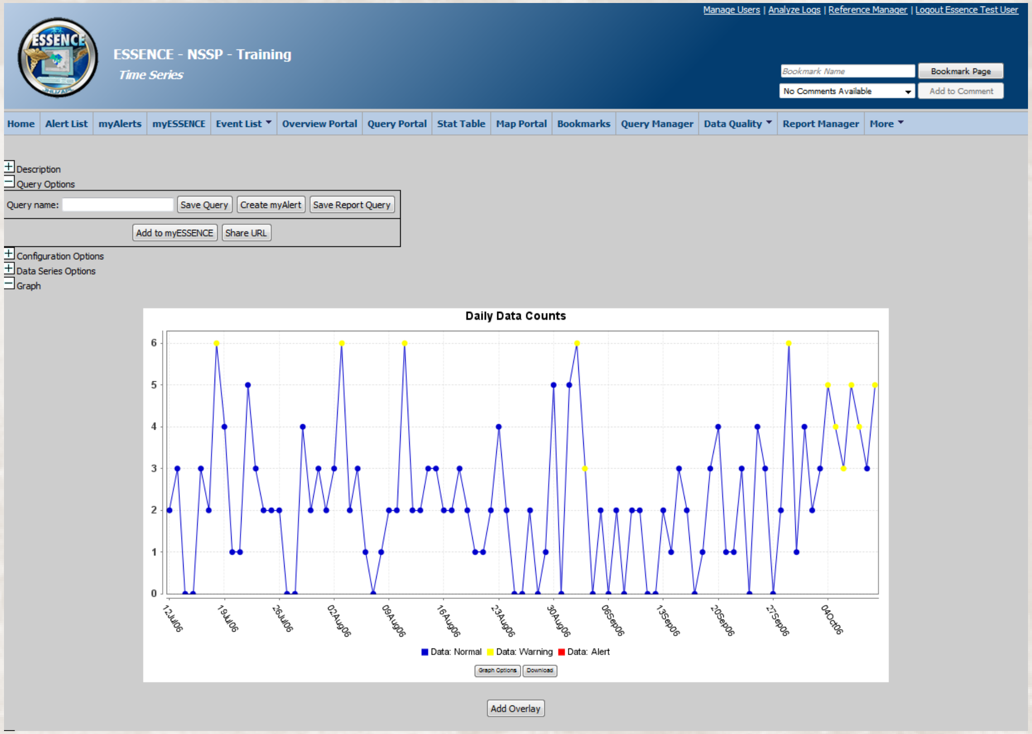




- **To perform free-text queries, choose the Chief Complaints parameter under the Medical Grouping System folder.**
- **The syntax for a chief complaint query is described in the help popup.**
- **Type in your free text query, then choose the select button to move it into your query definition.**

- **For the purpose of this walkthrough, please click on the Time Series button.**

Time Series Page

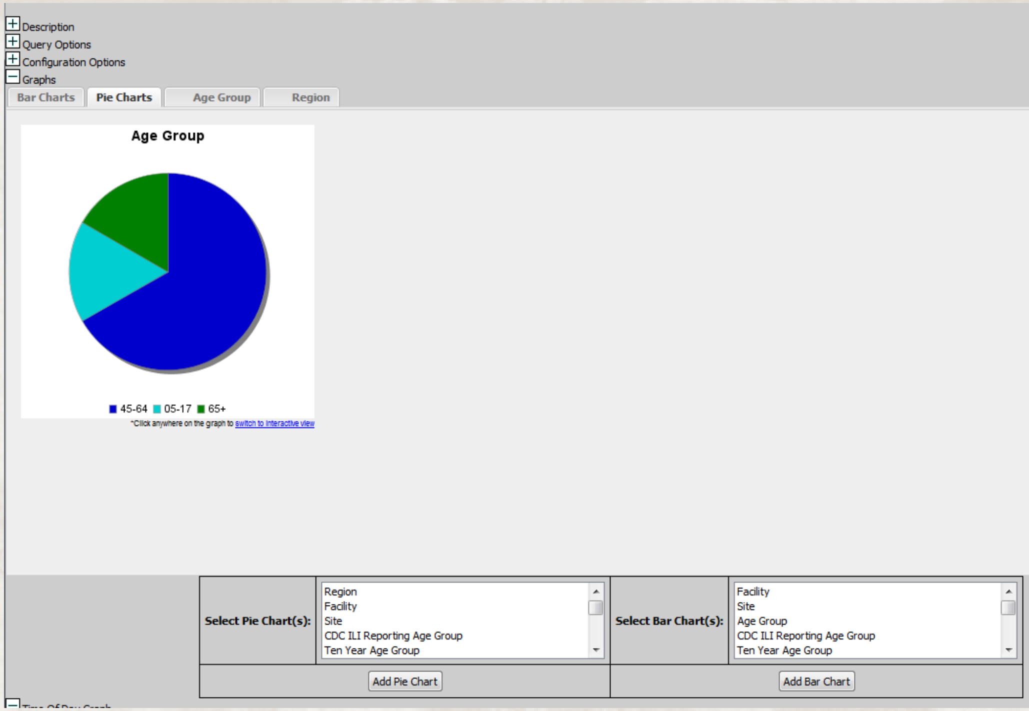


- A free-text query behaves just like any other query.

- For the purpose of this walkthrough, please click on a point on the graph to investigate the chief complaints in the Data Details page.

PIN	Date	Time	Facility Name	Zipcode	Region	C_Patient_County	ChiefComplaintOrig	ChiefComplaintParsed	Category_flat	S
Hosp08_NULL_VAER_7241	29Sep06	05:10 PM	Hospital 08	22101	VA_Fairfax	none	CHILLS/FEVER/BODY ACHES/COUGH	CHILLS FEVER BODY ACHES COUGH	;Resp;ILI;Fever;	;Chills;BodyAches;ILI;P_and_I;P_and
Hosp10_NULL_MDER_9118	29Sep06	02:11 AM	Hospital 10	OTHER	OTHER_REGION	none	CHILLS AND FEVER	CHILLS FEVER	;Fever;	;Chills;FeverOnly;FeverOrChills;Fever
Hosp12_NULL_MDER_2292	29Sep06	04:28 PM	Hospital 12	OTHER	OTHER_REGION	none	CHILL/FEVER COUGH	CHILL FEVER COUGH	;Resp;ILI;Fever;	;Chills;ILI;P_and_I;P_and_I_Without_
Hosp18_NULL_MDER_1610	29Sep06	03:26 AM	Hospital 18	OTHER	OTHER_REGION	none	CHILLS/FEVER/BODY ACHES/COUGH	CHILLS FEVER BODY ACHES COUGH	;Resp;ILI;Fever;	;Chills;BodyAches;ILI;P_and_I;P_and
Hosp20_NULL_MDER_8067	29Sep06	08:00 AM	Hospital 20	20852	MD_Montgomery	none	Fever/chills/sob	FEVER CHILLS SHORTNESS OF BREATH	;Resp;Fever;	;Chills;FeverOrChills;FeverPlus;Short
Hosp31_NULL_VACER2579	29Sep06	05:42 PM	Hospital 31	22180	VA_Fairfax	none	FEVER/CHILLS	FEVER CHILLS	;Fever;	;Chills;FeverOnly;FeverOrChills;Fever

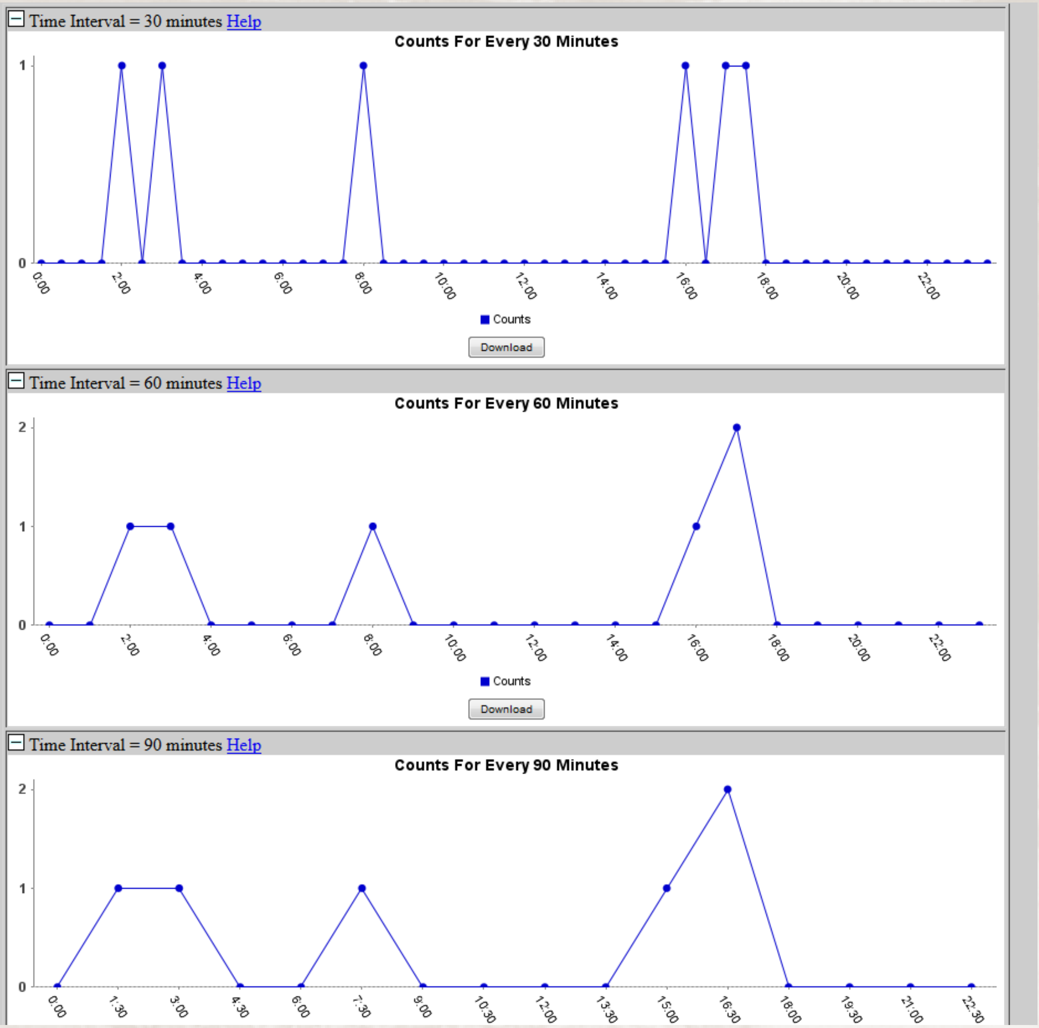
Data Details Page



- You can open up Pie and Bar charts for any parameter that has reference values.
- Additional tabs will be created with the data from the Pie / Bar chart.
- For the purpose of this walkthrough, please click on “Popup Time of Day Graphs” button.

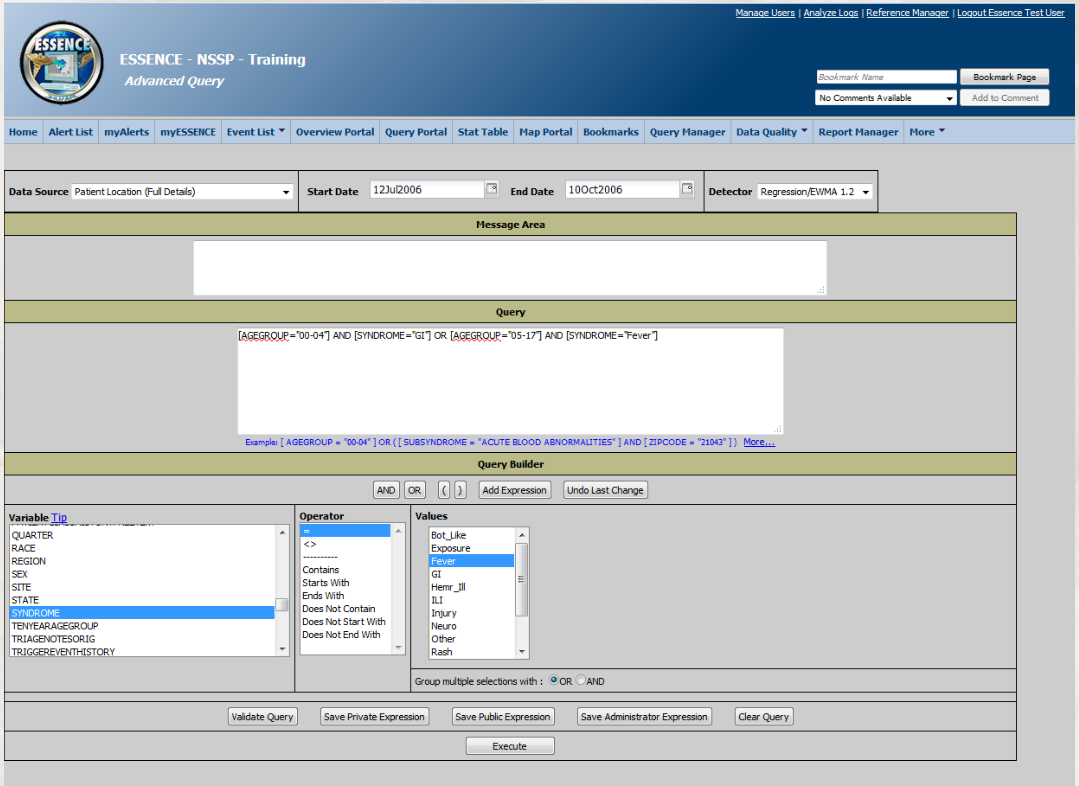
PIN	Date	Time	Facility Name	Zipcode	Region	C_Patient_County	ChiefComplaintOrig	ChiefComplaintParsed	Category_flat	S
Hosp08_NULL_VAER_7241	29Sep06	05:10 PM	Hospital 08	22101	VA_Fairfax	none	CHILLS/FEVER/BODY ACHES/COUGH	CHILLS FEVER BODY ACHES COUGH	;Resp;ILI;Fever;	;Chills;BodyAches;ILI;P_and_I;P_and
Hosp10_NULL_MDER_9118	29Sep06	02:11 AM	Hospital 10	OTHER	OTHER_REGION	none	CHILLS AND FEVER	CHILLS FEVER	;Fever;	;Chills;FeverOnly;FeverOrChills;Fever
Hosp12_NULL_MDER_2292	29Sep06	04:28 PM	Hospital 12	OTHER	OTHER_REGION	none	CHILL/FEVER COUGH	CHILL FEVER COUGH	;Resp;ILI;Fever;	;Chills;ILI;P_and_I;P_and_I_Without_
Hosp18_NULL_MDER_1610	29Sep06	03:26 AM	Hospital 18	OTHER	OTHER_REGION	none	CHILLS/FEVER/BODY ACHES/COUGH	CHILLS FEVER BODY ACHES COUGH	;Resp;ILI;Fever;	;Chills;BodyAches;ILI;P_and_I;P_and
Hosp20_NULL_MDER_8067	29Sep06	08:00 AM	Hospital 20	20852	MD_Montgomery	none	Fever/chills/sob	FEVER CHILLS SHORTNESS OF BREATH	;Resp;Fever;	;Chills;FeverOrChills;FeverPlus;Short
Hosp31_NULL_VACER2579	29Sep06	05:42 PM	Hospital 31	22180	VA_Fairfax	none	FEVER/CHILLS	FEVER CHILLS	;Fever;	;Chills;FeverOnly;FeverOrChills;Fever

Data Details Page



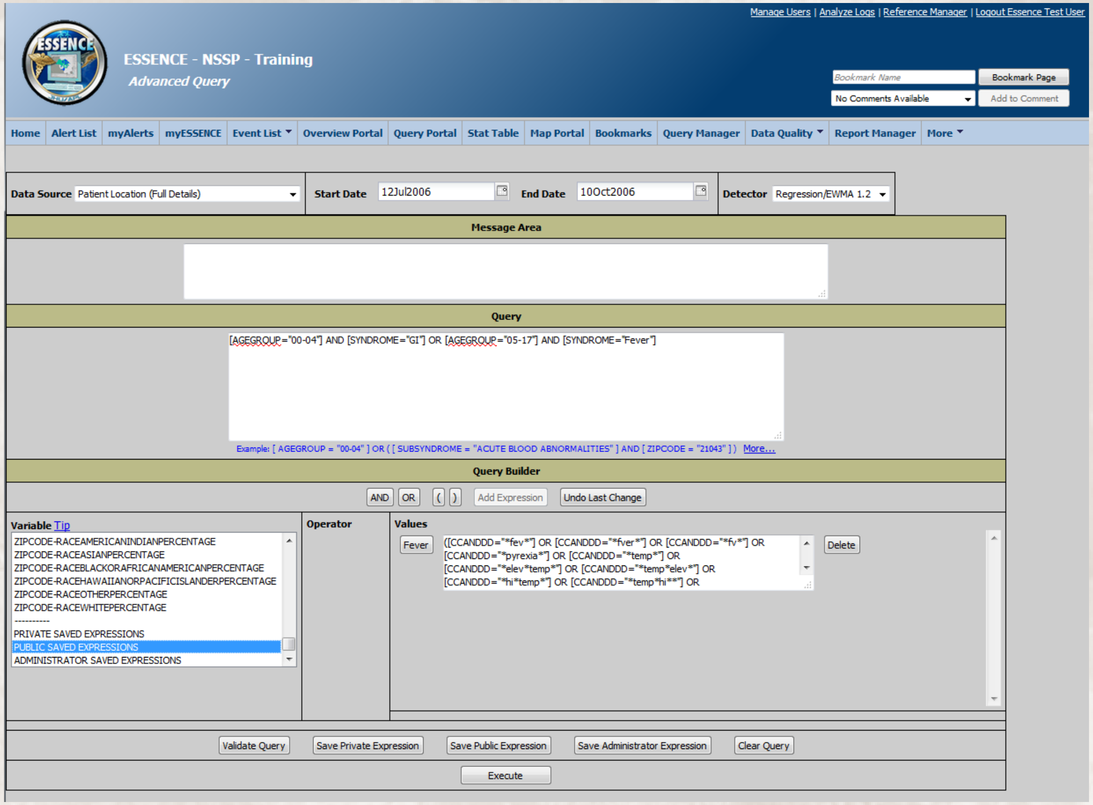
- You can view the data based on the Time of Arrival.
- For the purpose of this walkthrough, please click on the Back button on your browser, then click on the Query Portal.

Query Portal: AQT



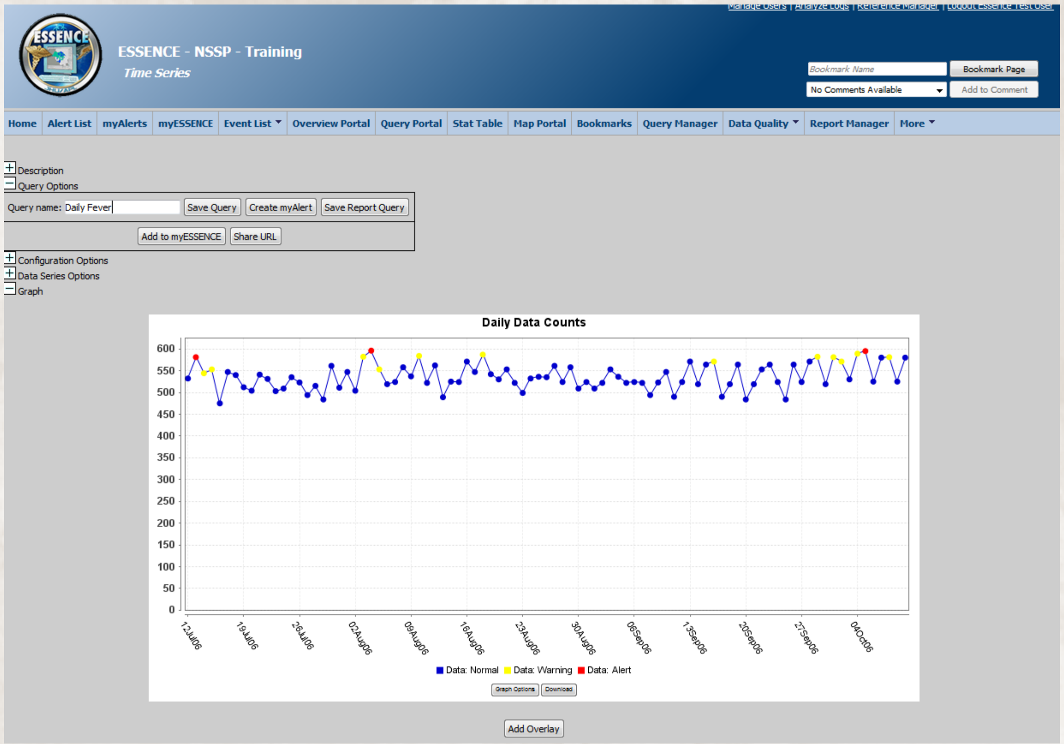
- For the purpose of this walkthrough, please choose the Adv Qry button
- The AQT screen allows you to create very complex queries.
- You can use the forms at the bottom to choose Variables, Operators, and Values.
- Once chosen, you can click “Add Expression” to put the expression into the Query window.
- You can also type your query directly into the Query Window.
- Continue on next slide…

Query Portal: AQT



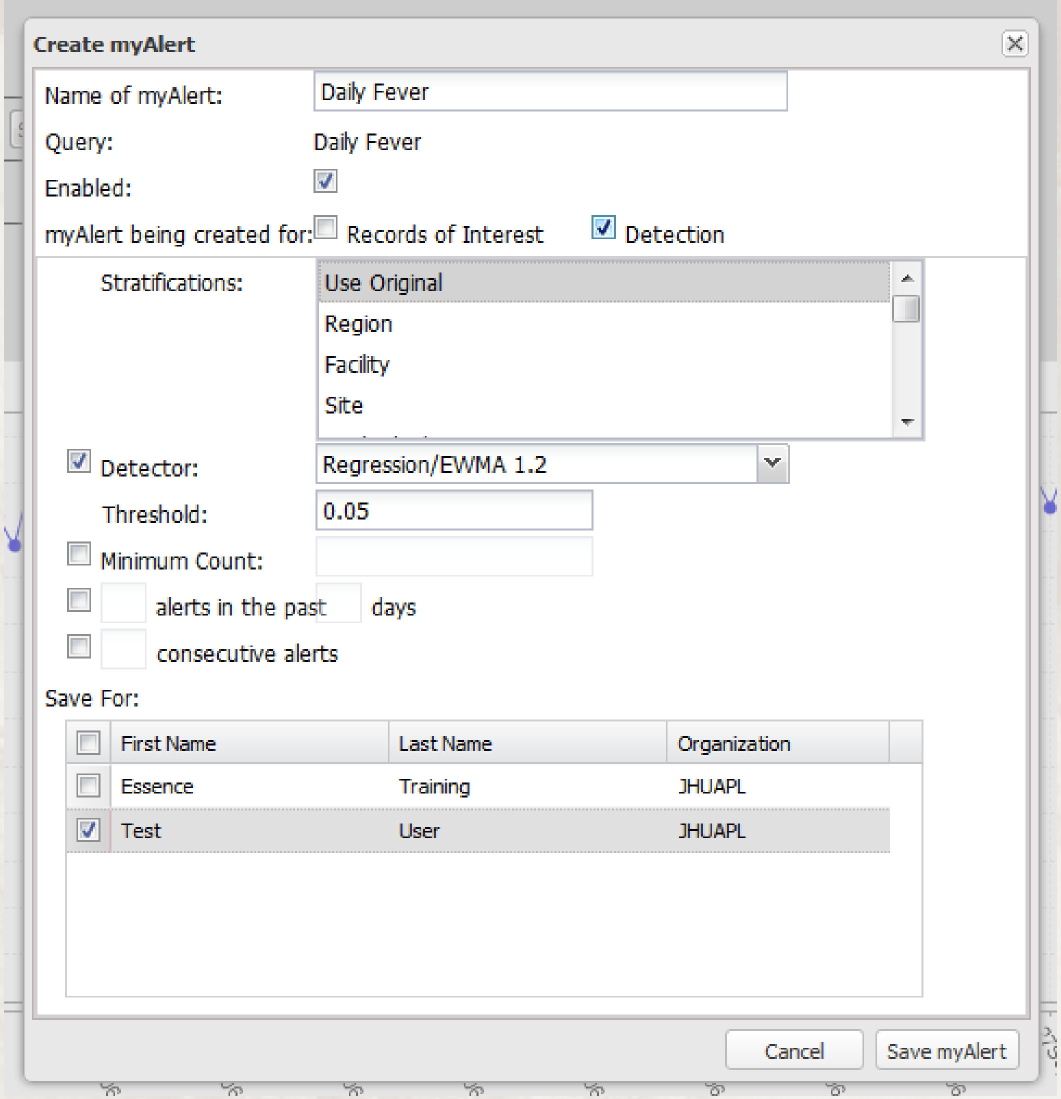
- You can save your expression privately with the “Save Private Expression” or publicly with the “Save Public Expression”.
- In the bottom of the Variable list, you can choose Private, Public, and Administrator Saved Expressions.
- Once chosen, you can click on the button of the expression and it will be added to your Query.
- Once you choose the Execute button, your query will be performed as a Time Series.
- For the purpose of this walkthrough, please click on the Query Portal.

Time Series: myAlerts



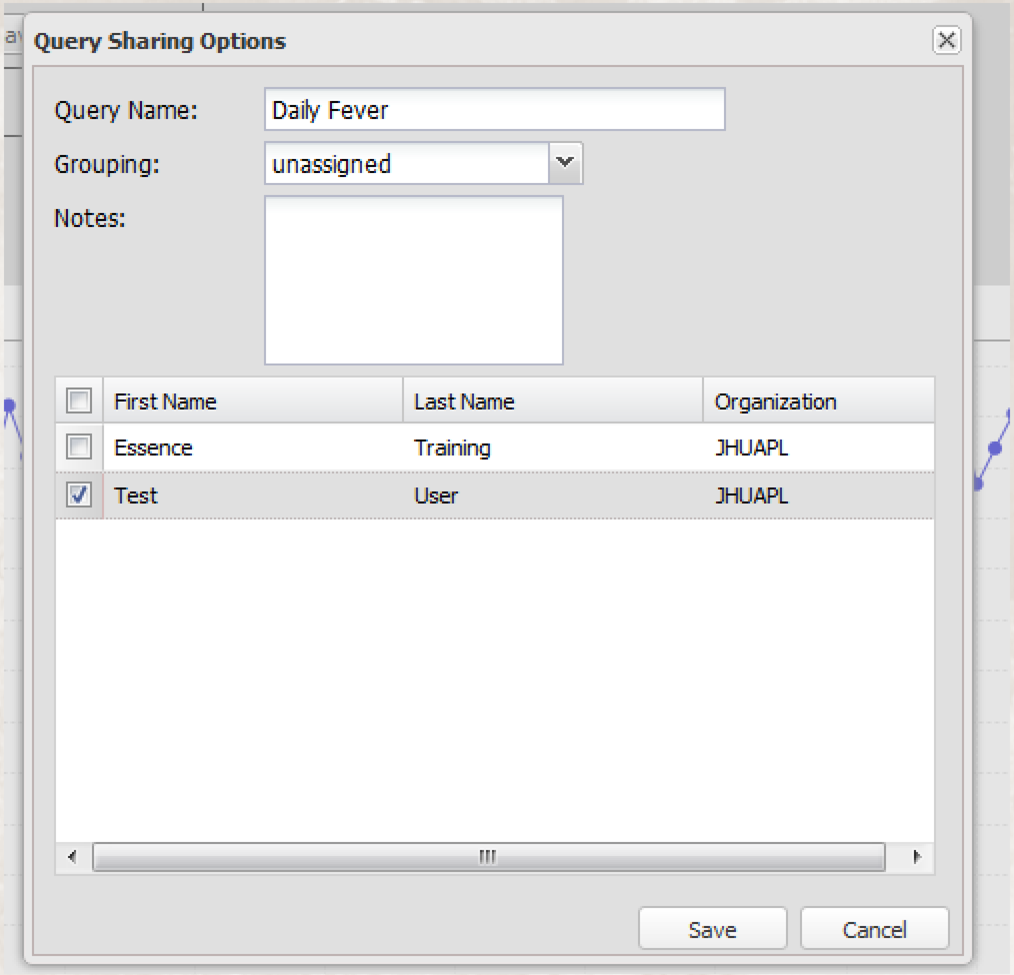
- Perform a Fever query, and view the Time Series of that query.
- In the Query Options section, you can name a query.
- Once named, a query can be Saved, used to create a myAlert, used to create a Report Query, added to a myESSENCE dashboard.
- For the purpose of this walkthrough, please click on the Create myAlert button.

Time Series: myAlerts



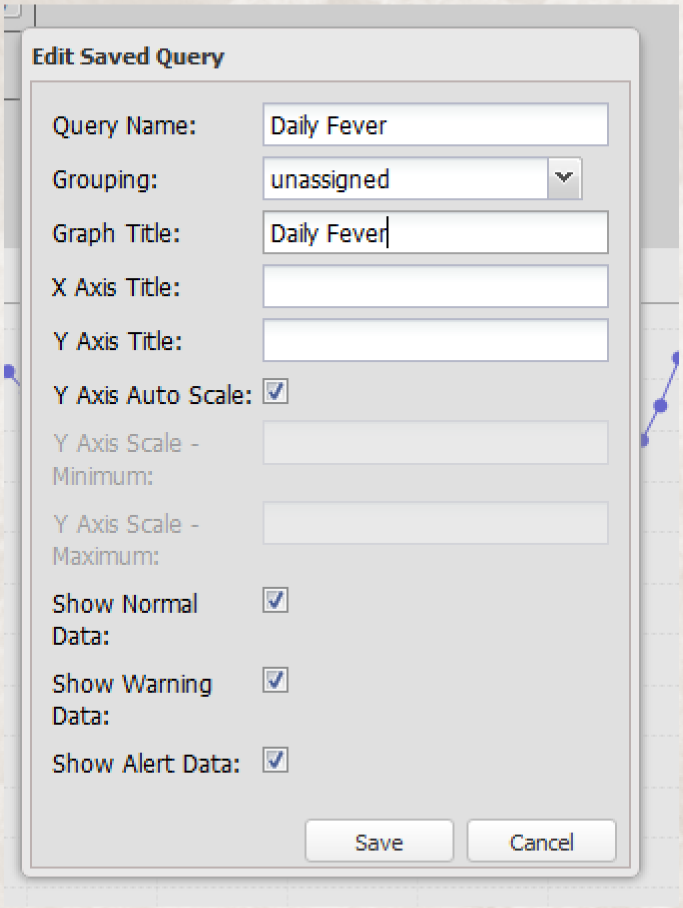
- **The "Records of Interest" option will create an myAlert for any record that meets the query definition.**
- **The "Detection" option allow you to determine the aspects of the detector you want.**
- **You can choose Detector and/or Minimum Count, but you must choose one.**
- **You can save a myAlert definition just for yourself or for multiple ESSENCE users.**
- **Saved myAlerts will run based on the back-end schedule for detectors. Results will not be available immediately.**
- **Cancel the myAlert creation, and continue to next slide…**

Time Series: Saved Queries



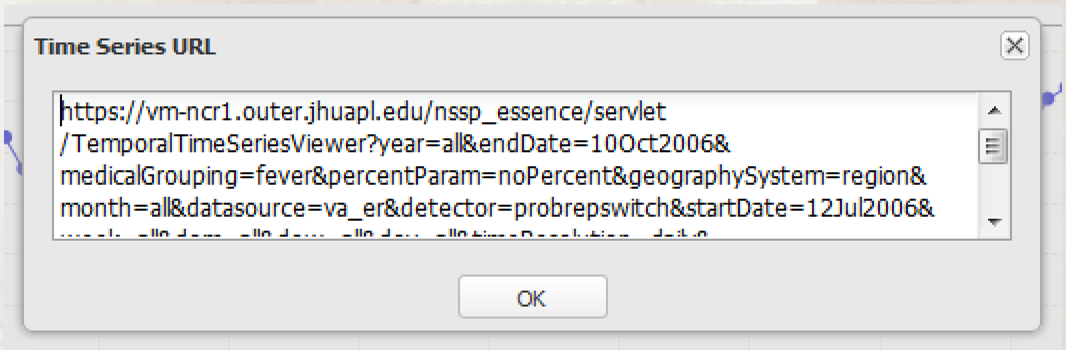
- The “Save Query” option will popup the window shown here.
- You can type in a new Grouping name if you want to organize your saved queries by name.
- Notes provide a place to describe your saved query, this is useful if sharing
- Can create the saved query for you or another ESSENCE user.
- For the purpose of this walkthrough, please click on the Save button.

Time Series: Report Saved Queries



- **The "Save Report Query" option will popup the window shown here.**
- **You can type in a Grouping name if you want to organize your saved queries by name.**
- **Report Queries are used in the MS Word Report System that will be explored later in this presentation.**
- **For the purpose of this walkthrough, please click on the Save button.**

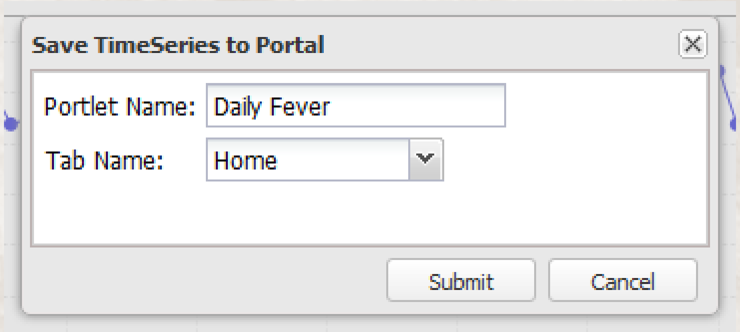
Query Portal: URL Sharing



- **The "Share URL" option will popup the window shown here.**
- **You can copy the URL and use it to email or send to others.**
- **This is done because if URLs are too long, the URL on the browser will not contain the information needed to recreate the query.**

- **For the purpose of this walkthrough, please click on the OK button.**

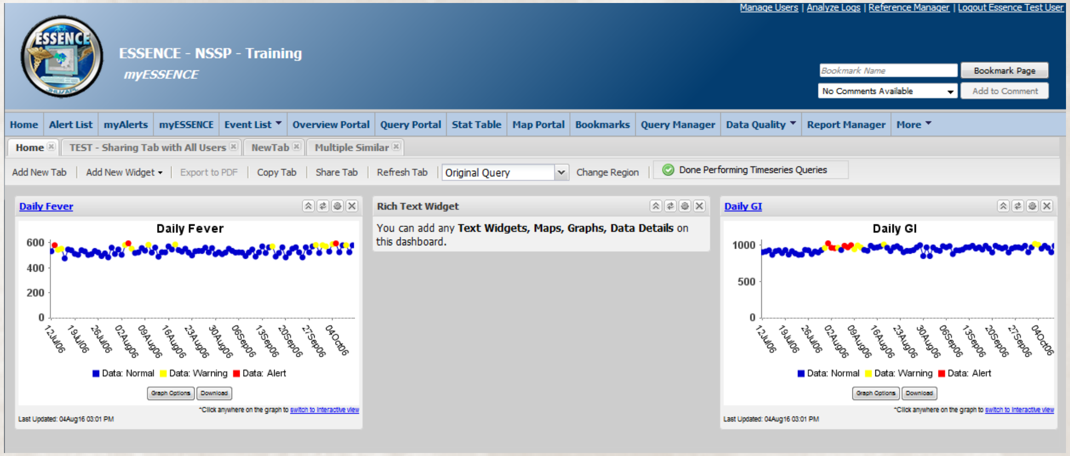
myESSENCE

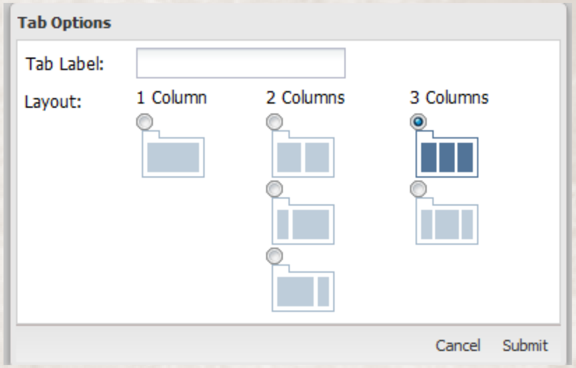


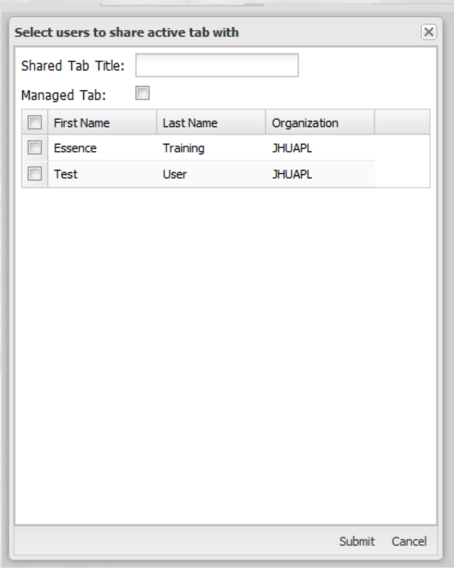
- The "Add to myESSENCE" option will popup the window shown here.
- You can name the graph to be added to your myESSENCE tab.
- You can choose which myESSENCE tab the graph is added to.

- For the purpose of this walkthrough, please click on the Submit button. Then click on the myESSENCE option from the main ESSENCE menu bar.

myESSENCE







- You can create new tabs.
- You can add widgets (easier to do it from Time Series, Data Details, Overview pages)
- Copy / Share Tab
- Sharing can be done by giving a copy to another user or “Managed” sharing, which shares a read-only version that you remain in control of.
- Filter to change the geography of most graphs (depends on data source).
- Can drag-n-drop widgets to re-organize them.

- For the purpose of this walkthrough, please click on the myAlert option from the main ESSENCE menu bar.

myAlerts

Alerts Messages

Records of Interest Messages

Manage Alert Definitions   Subscribe

Alerts | Records of Interest

Alert Definition	Stratifications	Date	Data Source	Level	Count	Expected	Timeseries
Daily Fever	Use Original	04Oct06	ER by Patient	0.005	420	363.36	Timeseries
Daily Fever	Use Original	05Oct06	ER by Patient	0.001	424	363.96	Timeseries
Daily Hosp Fever w Fairfax Patients	Use Original	04Oct06	ER by Hospital	0.049	70	60.50	Timeseries
Daily Fever	Use Original	07Oct06	ER by Patient	0.013	415	366.39	Timeseries
Daily Hosp Fever w Fairfax Patients	Use Original	05Oct06	ER by Hospital	0.006	78	60.14	Timeseries
Daily Fever	Use Original	08Oct06	ER by Patient	0.006	419	368.32	Timeseries
Daily Hosp Fever w Fairfax Patients	Use Original	06Oct06	ER by Hospital	0.006	75	60.36	Timeseries
Daily Hosp Fever w Fairfax Patients	Use Original	07Oct06	ER by Hospital	0.032	67	60.86	Timeseries
Daily Fever	Use Original	10Oct06	ER by Patient	0.029	415	371.36	Timeseries
Daily Hosp Fever w Fairfax Patients	Use Original	08Oct06	ER by Hospital	0.014	76	61.14	Timeseries
Daily Hosp Fever w Fairfax Patients	Use Original	09Oct06	ER by Hospital	0.014	75	61.86	Timeseries
Daily Hosp Fever w Fairfax Patients	Use Original	10Oct06	ER by Hospital	0.047	67	62.29	Timeseries
Daily Hosp Fever w Fairfax Patients	Age Group: 5-17	04Oct06	ER by Hospital	0.043	15	10.82	Timeseries
Daily Hosp Fever w Fairfax Patients	Age Group: 5-17	05Oct06	ER by Hospital	0.001	20	10.86	Timeseries
Daily Hosp Fever w Fairfax Patients	Age Group: 5-17	08Oct06	ER by Hospital	0.015	19	11.61	Timeseries
Daily Hosp Fever w Fairfax Patients	Age Group: 0-4	04Oct06	ER by Hospital	0.025	38	30.07	Timeseries
Daily Hosp Fever w Fairfax Patients	Age Group: 0-4	06Oct06	ER by Hospital	0.042	35	29.93	Timeseries
Daily Hosp Fever w Fairfax Patients	Age Group: 0-4	07Oct06	ER by Hospital	0.012	38	30.29	Timeseries
Daily Hosp Fever w Fairfax Patients	Age Group: 0-4	08Oct06	ER by Hospital	0.049	30	29.96	Timeseries
Daily Hosp Fever w Fairfax Patients	Age Group: 0-4	09Oct06	ER by Hospital	0.032	35	30.25	Timeseries
Daily Hosp Fever w Fairfax Patients	Age Group: 0-4	10Oct06	ER by Hospital	0.009	38	30.43	Timeseries
Daily AQT Fever	Use Original	04Oct06	ER by Patient	0.005	420	363.36	Timeseries

Alerts Messages

Records of Interest Messages

Manage Alert Definitions   Subscribe

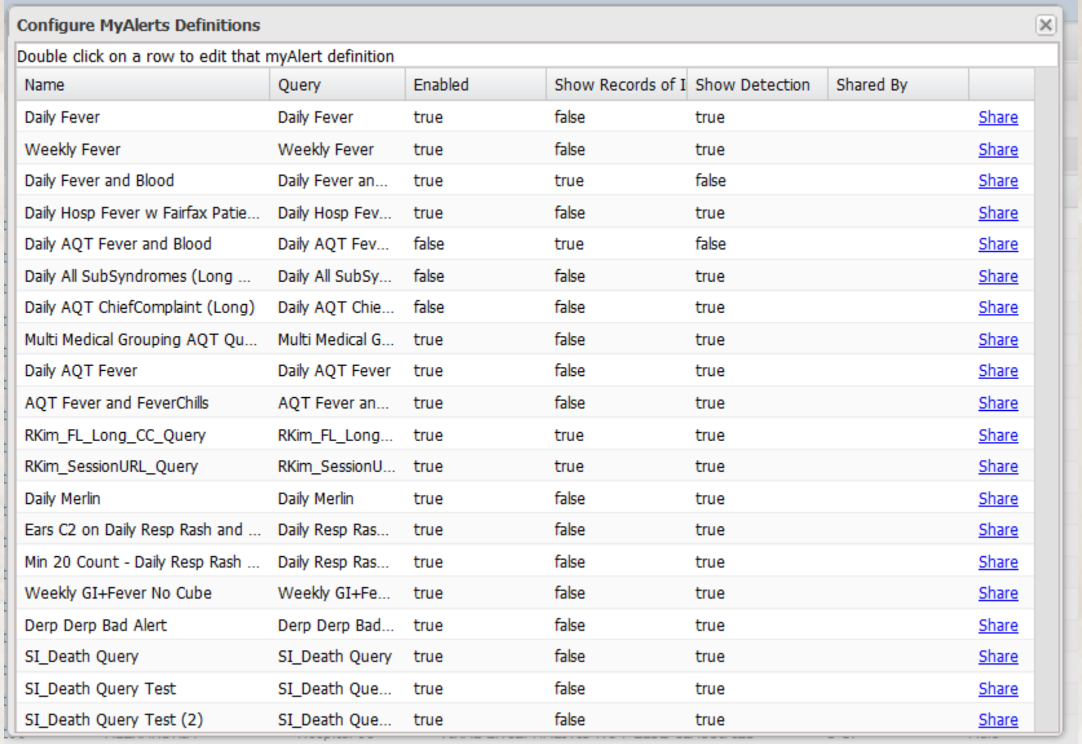
Alerts | Records of Interest

Alert Definition	Date	Geography	Facility	Medical Grouping	Age Group	Sex	Data Details
Daily Fever and Blood	09Oct06	PRINCE GEORGES	Hospital 19	CHEST PAIN FEVER BLOOD VMT	18-44	Female	Data Details
Daily Fever and Blood	06Oct06	PRINCE GEORGES	Hospital 19	CHEST PAIN FEVER BLOOD VMT	18-44	Female	Data Details
Daily Fever and Blood	03Oct06	PRINCE GEORGES	Hospital 19	CHEST PAIN FEVER BLOOD VMT	18-44	Female	Data Details
Daily Fever and Blood	10Oct06	OTHER_REGION	Hospital 35	FEVER LOW BLOOD PRESSURE	45-64	Female	Data Details
Daily Fever and Blood	07Oct06	OTHER_REGION	Hospital 35	FEVER LOW BLOOD PRESSURE	45-64	Female	Data Details
Daily Fever and Blood	04Oct06	OTHER_REGION	Hospital 35	FEVER LOW BLOOD PRESSURE	45-64	Female	Data Details
Daily Fever and Blood	08Oct06	PRINCE WILLIAM	Hospital 38	COUGHING BLOOD FEVER	5-17	Female	Data Details
Daily Fever and Blood	05Oct06	PRINCE WILLIAM	Hospital 38	COUGHING BLOOD FEVER	5-17	Female	Data Details
Daily Fever and Blood	05Oct06	WASHINGTON	Hospital 42	FEVER SPITTING BLOOD	65+	Male	Data Details
Daily Fever and Blood	05Oct06	WASHINGTON	Hospital 42	FEVER SPITTING BLOOD	65+	Male	Data Details
RKim_FL_Long_CC_Query	08Oct06	OTHER_REGION	Hospital 02	EVALUATION FOR RABIES	45-64	Male	Data Details
RKim_FL_Long_CC_Query	05Oct06	OTHER_REGION	Hospital 02	EVALUATION FOR RABIES	45-64	Male	Data Details
RKim_FL_Long_CC_Query	09Oct06	WASHINGTON	Hospital 05	MENINGITIS	18-44	Female	Data Details
RKim_FL_Long_CC_Query	06Oct06	WASHINGTON	Hospital 05	MENINGITIS	18-44	Female	Data Details
RKim_FL_Long_CC_Query	03Oct06	WASHINGTON	Hospital 05	MENINGITIS	18-44	Female	Data Details
RKim_FL_Long_CC_Query	08Oct06	ALEXANDRIA	Hospital 08	VIRAL ENCEPHALITIS NOT ELSE CLASSIFIED	5-17	Male	Data Details
RKim_FL_Long_CC_Query	05Oct06	ALEXANDRIA	Hospital 08	VIRAL ENCEPHALITIS NOT ELSE CLASSIFIED	5-17	Male	Data Details
RKim_FL_Long_CC_Query	10Oct06	OTHER_REGION	Hospital 12	MENINGITIS NOT OTHERWISE SPECIFIED	5-17	Male	Data Details
RKim_FL_Long_CC_Query	07Oct06	OTHER_REGION	Hospital 12	MENINGITIS NOT OTHERWISE SPECIFIED	5-17	Male	Data Details
RKim_FL_Long_CC_Query	04Oct06	OTHER_REGION	Hospital 12	MENINGITIS NOT OTHERWISE SPECIFIED	5-17	Male	Data Details
RKim_FL_Long_CC_Query	08Oct06	MONTGOMERY	Hospital 13	EXPOSURE TO RABIES	18-44	Female	Data Details
RKim_FL_Long_CC_Query	08Oct06	PRINCE GEORGES	Hospital 13	RULE OUT MENINGITIS	18-44	Female	Data Details

- **When myAlerts are created by the back-end process you can view Alerts and Records of Interest.**

- **Continue on next slide…**

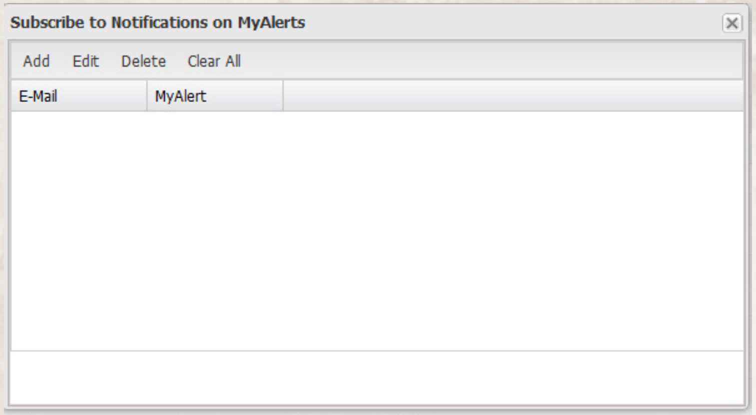
myAlerts



Configure MyAlerts Definitions

Double click on a row to edit that myAlert definition

Name	Query	Enabled	Show Records of I	Show Detection	Shared By	
Daily Fever	Daily Fever	true	false	true		Share
Weekly Fever	Weekly Fever	true	false	true		Share
Daily Fever and Blood	Daily Fever an...	true	true	false		Share
Daily Hosp Fever w Fairfax Patie...	Daily Hosp Fev...	true	false	true		Share
Daily AQT Fever and Blood	Daily AQT Fev...	false	true	false		Share
Daily All SubSyndromes (Long ...	Daily All SubSy...	false	false	true		Share
Daily AQT ChiefComplaint (Long)	Daily AQT Chie...	false	false	true		Share
Multi Medical Grouping AQT Qu...	Multi Medical G...	true	false	true		Share
Daily AQT Fever	Daily AQT Fever	true	false	true		Share
AQT Fever and FeverChills	AQT Fever an...	true	false	true		Share
RKim_FL_Long_CC_Query	RKim_FL_Long...	true	true	true		Share
RKim_SessionURL_Query	RKim_SessionU...	true	true	true		Share
Daily Merlin	Daily Merlin	true	false	true		Share
Ears C2 on Daily Resp Rash and ...	Daily Resp Ras...	true	false	true		Share
Min 20 Count - Daily Resp Rash ...	Daily Resp Ras...	true	false	true		Share
Weekly GI+Fever No Cube	Weekly GI+Fe...	true	false	true		Share
Derp Derp Bad Alert	Derp Derp Bad...	true	false	true		Share
SI_Death Query	SI_Death Query	true	false	true		Share
SI_Death Query Test	SI_Death Que...	true	false	true		Share
SI_Death Query Test (2)	SI_Death Que...	true	false	true		Share



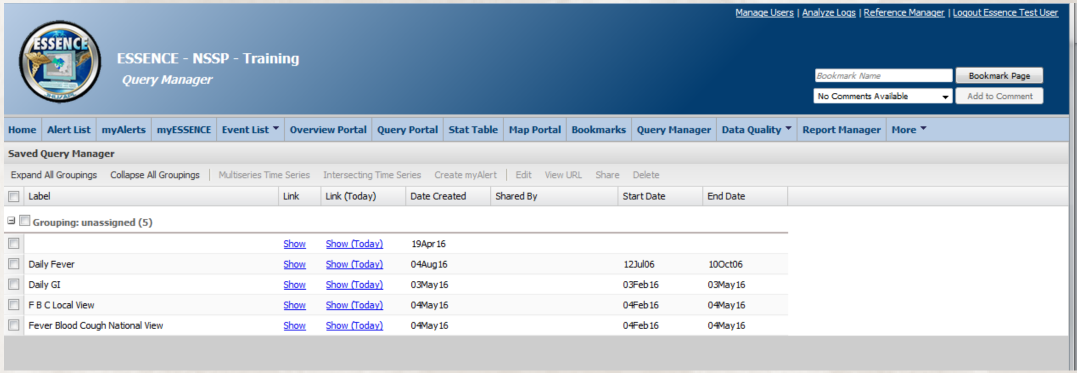
Subscribe to Notifications on MyAlerts

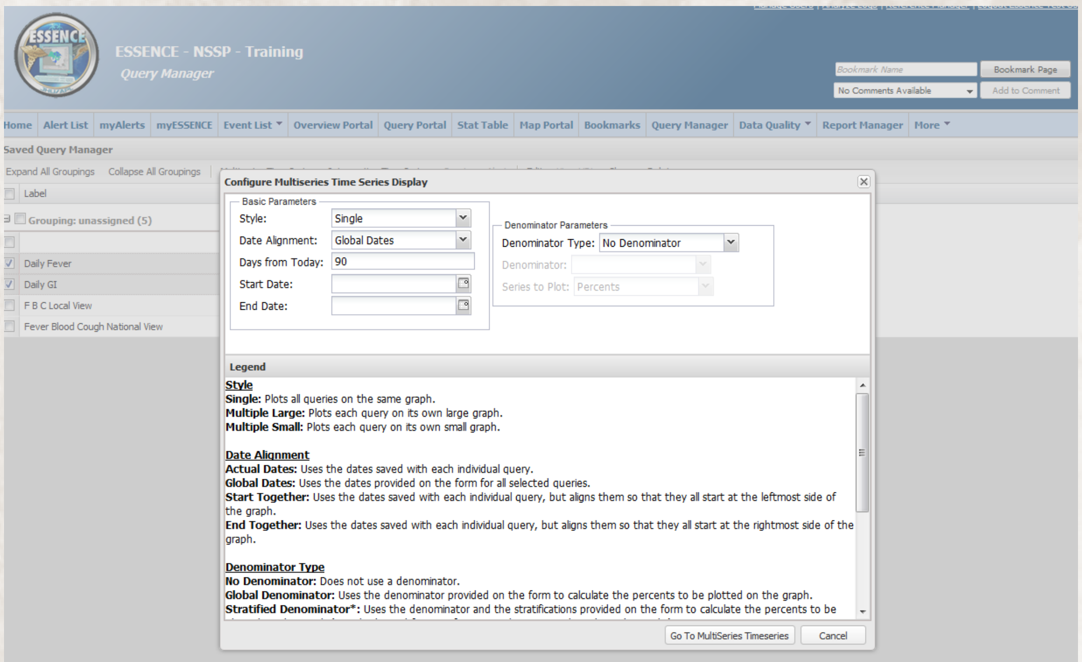
Add Edit Delete Clear All

E-Mail	MyAlert

- **The Manage Alert Definitions option pops up the window shown here.**
- **You can double click on a definition to edit it.**
- **The Subscribe option allows you to setup email subscriptions for myAlerts.**
- **For the purpose of this walkthrough, please click on the Query Manager option from the main ESSENCE menu bar.**

Query Manager

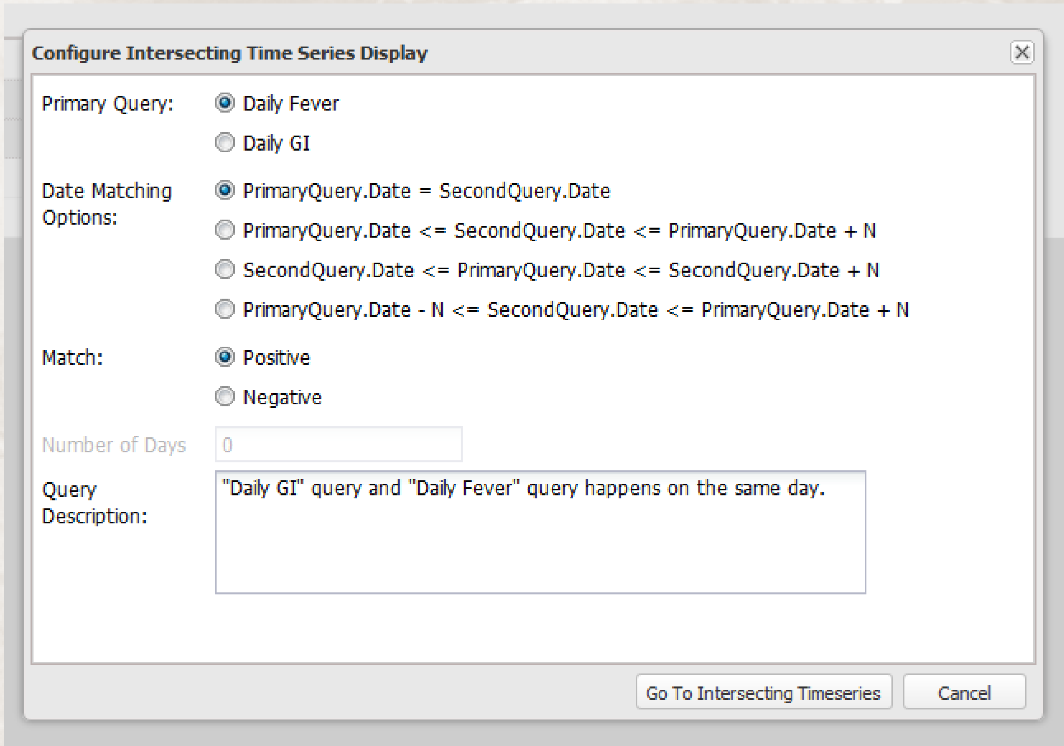




- Saved Queries can be viewed as they were originally saved (Show) or with the start date end date shifted so that the end date = today using the Show (Today) link.
- If you choose multiple saved queries, you can create a Multi-Series Time Series Graph

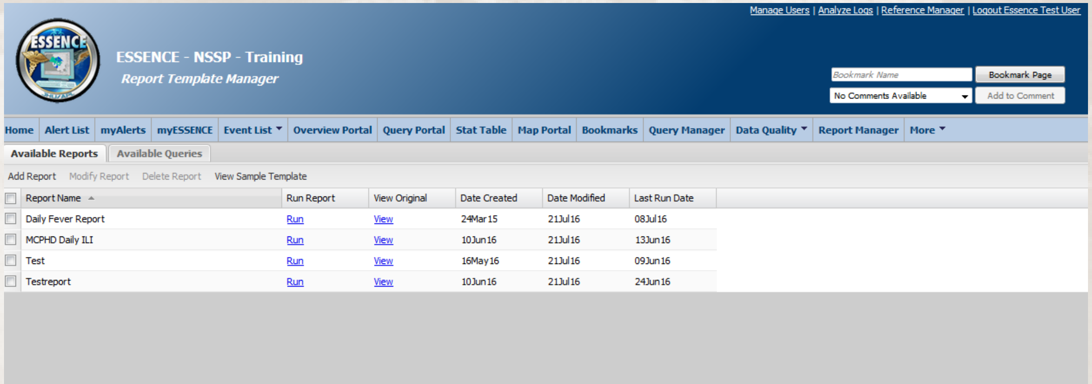
- Continue on next slide…

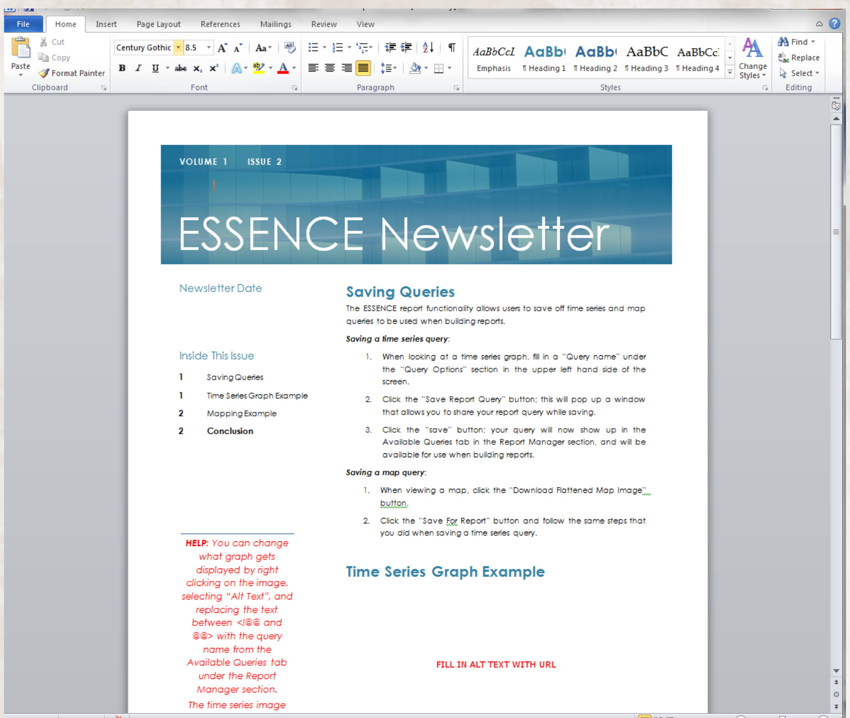
Query Manager



- **Intersecting Time Series takes two queries and finds all records that positively or negatively match between the two queries.**
- **For the purpose of this walkthrough, please click on the Report Manager option from the main ESSENCE menu bar.**

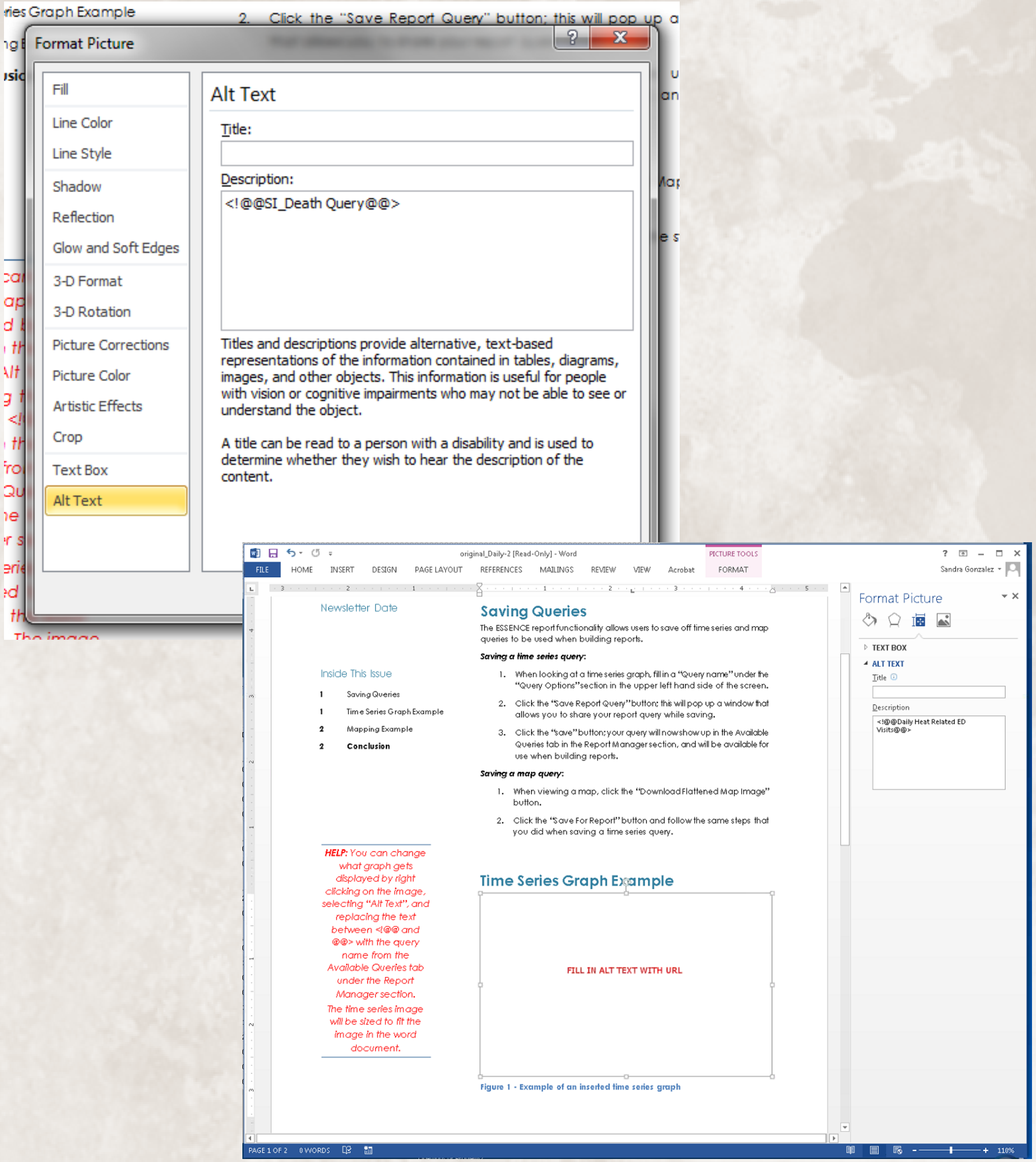
Report Manager





- **By Viewing the Sample Template, a MS Word document will be downloaded.**
- **The sample contains instructions on how to edit / save a new report.**
- **For the purpose of this walkthrough, download the sample.**

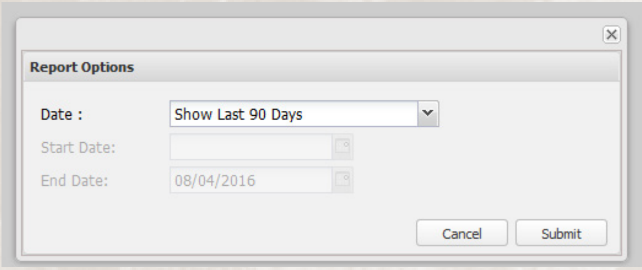
Report Manager

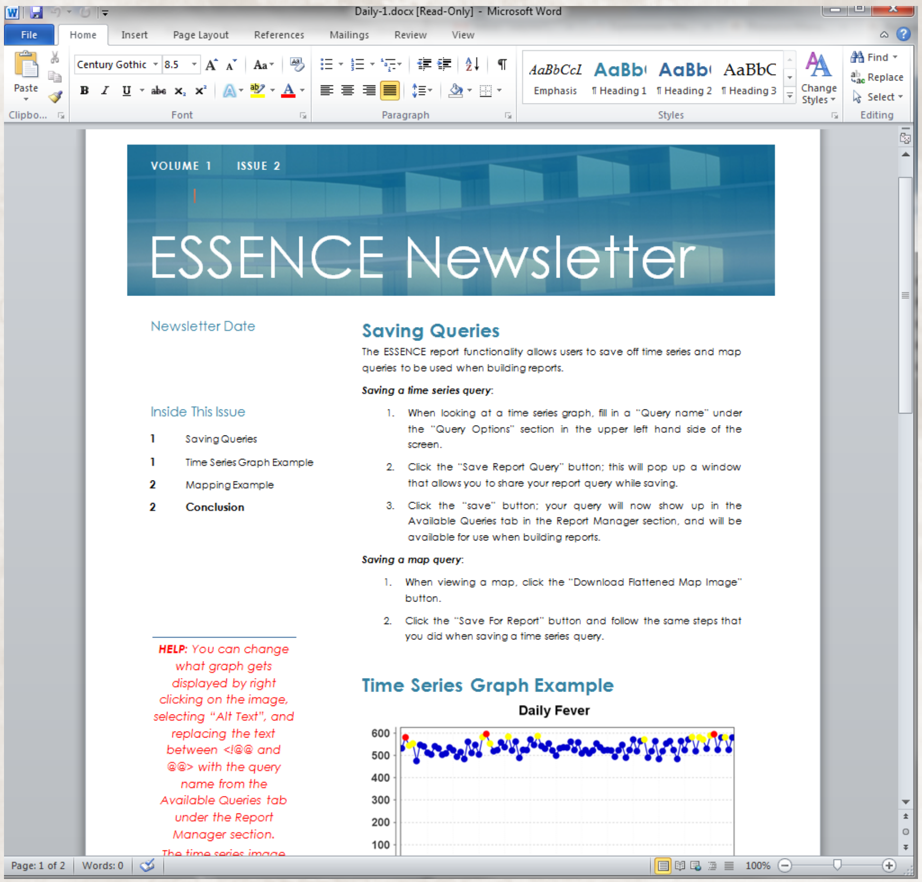


- Right-Click on the image and select the Format Picture…
- In the Alt Text section, replace the SI_Death Query with the name of the query you want embedded.
- The saved MS Word document can then be uploaded as a new report.

- For the purpose of this walkthrough, do not upload a new report, just click Run on an existing report.

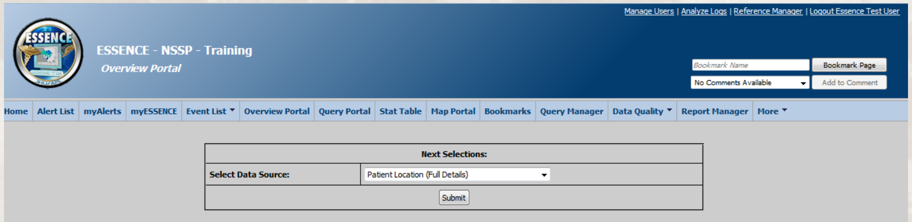
Report Manager

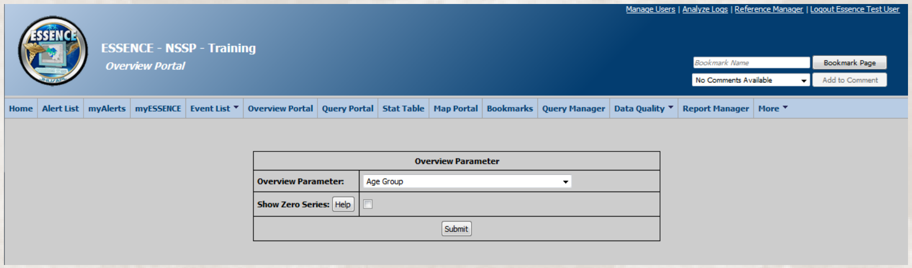




- You can choose the date range you want, then submit to run the report.
- A MS Word document will be created with the embedded graphs or maps in the document.
- For the purpose of this walkthrough, please click on the Overview Portal option from the main ESSENCE menu bar.

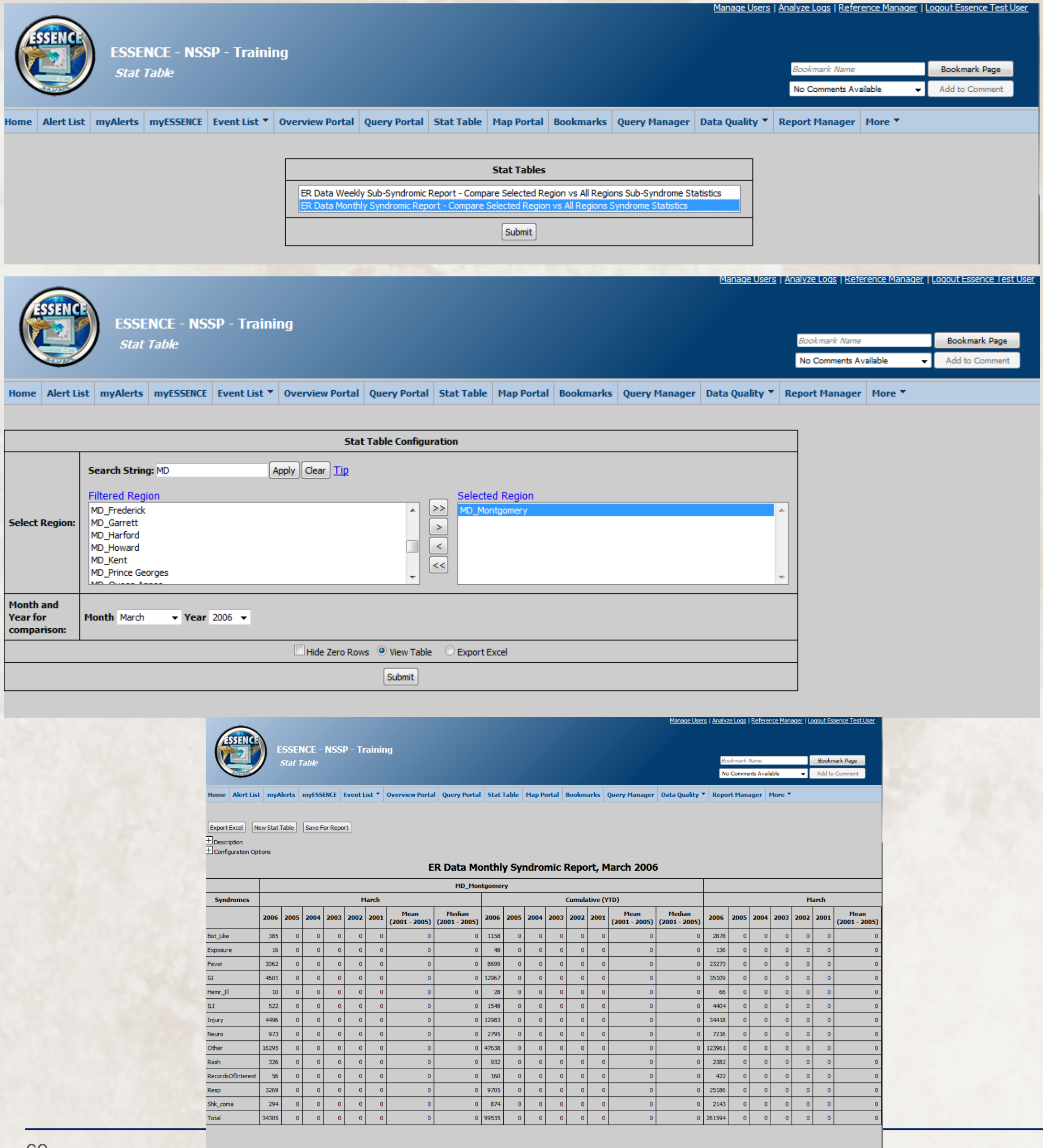
Overview Portal





- The Overview Portal can be accessed two ways: the Overview Portal menu option or from a Query Wizard.
- If you enter the Overview Portal from the menu button, you will get the default options for the datasource you choose.
- If you enter from the Query Wizard, you can choose the parameters you want pre-defined before entering the overview portal.
- The functionality of the Overview Portal has been almost entirely replaced by the Stratification system on the Time Series Page.
- The last remaining feature that has not been duplicated is the ability to add all the overview graphs to a myESSENCE dashboard with a single click.
- If you wish to perform an overview by hospital or region – it is best to down select those in a Query Portal first, to minimize the amount of querying the system must do to create graphs for every region or every hospital across the entire country.
- You can also download a zip file containing all the graphs from the link at the bottom of the page.
- For the purpose of this walkthrough, please click on the Stat Table option from the main ESSENCE menu bar.

Stat Table



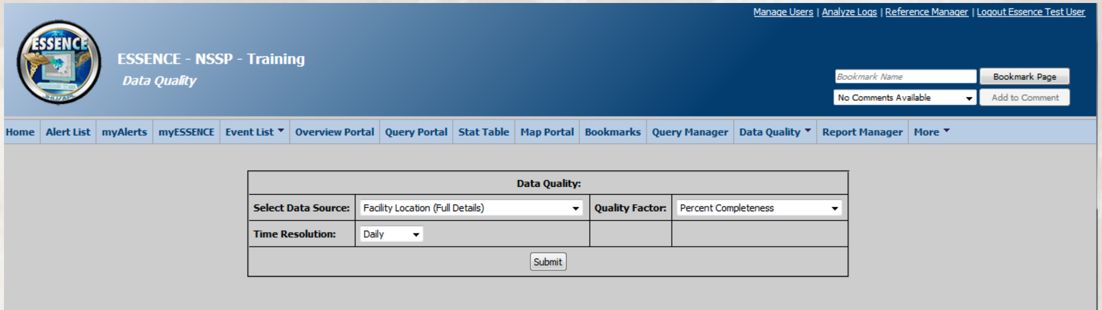
ER Data Monthly Syndromic Report, March 2006

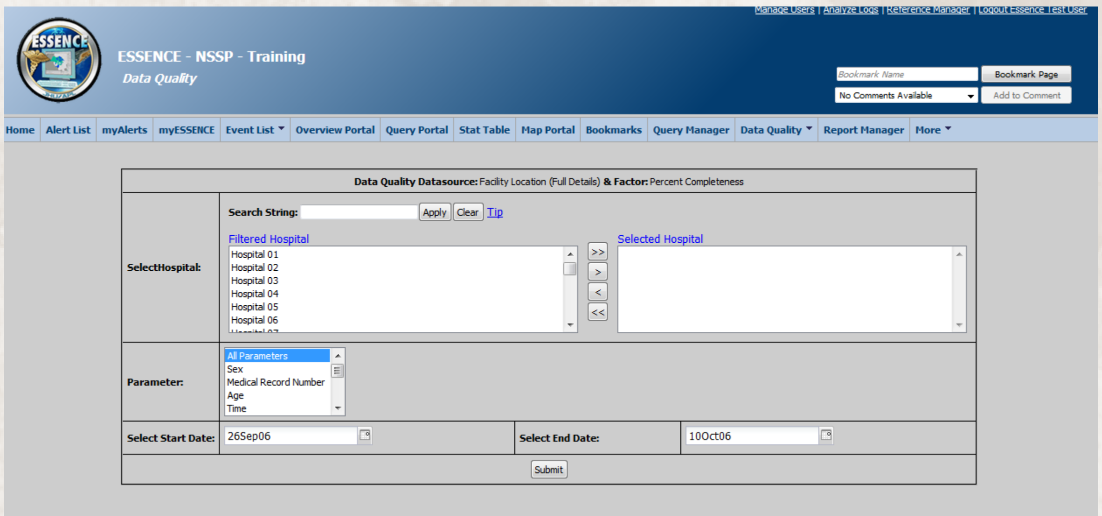
	MD_Montgomery																						
Syndromes	March								Cumulative (YTD)								March						
	2006	2005	2004	2003	2002	2001	Mean (2001 - 2005)	Median (2001 - 2005)	2006	2005	2004	2003	2002	2001	Mean (2001 - 2005)	Median (2001 - 2005)	2006	2005	2004	2003	2002	2001	Mean (2001 - 2005)
Bot_Like	385	0	0	0	0	0	0	0	1158	0	0	0	0	0	0	0	2878	0	0	0	0	0	0
Exposure	16	0	0	0	0	0	0	0	48	0	0	0	0	0	0	0	136	0	0	0	0	0	0
Fever	3062	0	0	0	0	0	0	0	8699	0	0	0	0	0	0	0	23273	0	0	0	0	0	0
GI	4601	0	0	0	0	0	0	0	12967	0	0	0	0	0	0	0	35109	0	0	0	0	0	0
Hemr_Ill	10	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0	66	0	0	0	0	0	0
ILI	522	0	0	0	0	0	0	0	1548	0	0	0	0	0	0	0	4404	0	0	0	0	0	0
Injury	4496	0	0	0	0	0	0	0	12983	0	0	0	0	0	0	0	34418	0	0	0	0	0	0
Neuro	973	0	0	0	0	0	0	0	2795	0	0	0	0	0	0	0	7216	0	0	0	0	0	0
Other	16295	0	0	0	0	0	0	0	47638	0	0	0	0	0	0	0	123961	0	0	0	0	0	0
Rash	326	0	0	0	0	0	0	0	932	0	0	0	0	0	0	0	2382	0	0	0	0	0	0
RecordsOfInterest	56	0	0	0	0	0	0	0	160	0	0	0	0	0	0	0	422	0	0	0	0	0	0
Resp	3269	0	0	0	0	0	0	0	9705	0	0	0	0	0	0	0	25186	0	0	0	0	0	0
Shk_coma	294	0	0	0	0	0	0	0	874	0	0	0	0	0	0	0	2143	0	0	0	0	0	0
Total	34305	0	0	0	0	0	0	0	99535	0	0	0	0	0	0	0	261594	0	0	0	0	0	0

- **The Stat Table provides pre-built reporting capabilities.**
- **Choose a report, and complete the required form.**
- **The report will then be created and available for view in Excel or in the web page.**

- **For the purpose of this walkthrough, please click on the Data Quality option from the main ESSENCE menu bar.**

Data Quality





- **The Data Quality portal has a few different options.**
- **The first allows you to view the Percent Completeness, Percent Mapped to Known Values, and the Percent Received Within 24 Hours for any data source that has been Data Quality configured.**
- **You can choose specific facilities (recommended) or parameters to view.**

- **Continue on next slide…**

Data Quality

Legend (Percent %)						
96 - 100	91 - 95	81 - 90	51 - 80	06 - 50	00 - 05	N/A

Data Quality

Geography	Variable	26Sep06	27Sep06	28Sep06	29Sep06	30Sep06	01Oct06	02Oct06	03Oct06	04Oct06	05Oct06	06Oct06	07Oct06	08Oct06	09Oct06	10Oct06
Hospital 01	Sex	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Hospital 01	Medical Recor...	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Hospital 01	Age	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Hospital 01	Time	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Hospital 01	Zipcode	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Hospital 01	Discharge Dia...	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hospital 01	Discharge Dis...	7	7	8	7	4	8	8	5	11	8	5	11	8	5	11
Hospital 01	Mode of Arrival	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hospital 01	Chief Complaint	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Hospital 02	Sex	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Hospital 02	Medical Recor...	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Hospital 02	Age	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Hospital 02	Time	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Hospital 02	Zipcode	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Hospital 02	Discharge Dia...	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hospital 02	Discharge Dis...	9	7	7	10	5	7	7	9	8	7	9	8	7	9	8
Hospital 02	Mode of Arrival	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hospital 02	Chief Complaint	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

- **The results will be displayed in a color coded table.**
- **For the purpose of this walkthrough, please click on the Data Quality - Alerts option from the main ESSENCE menu bar.**

Data Quality

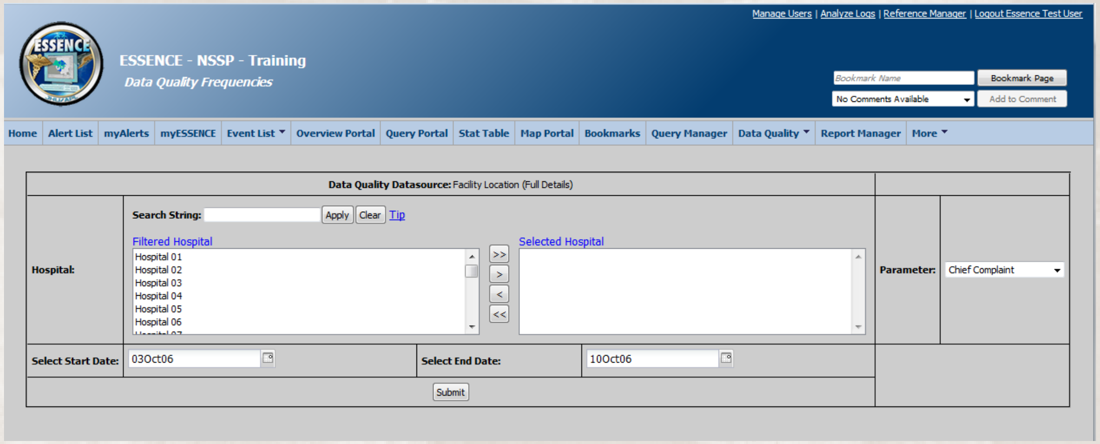
- **Data Quality Alerts will show any factor that has changed (+ / -) 10%.**

- **For the purpose of this walkthrough, please click on the Data Quality - Frequencies option from the main ESSENCE menu bar.**

Data Quality Alerts

Date	Data Sourc	Geography	Variable	Quality Factor	Percent	Previous Day	Percent Difference	Variable	Total Count
09Oct06	ER Da...	Hospi...	Discharge Disposition	Percent Completeness	13.16	2.63	10.53	5	38
09Oct06	ER Da...	Hospi...	Discharge Disposition	Percent Mapped to Known Values	13.16	2.63	10.53	5	38
06Oct06	ER Da...	Hospi...	Discharge Disposition	Percent Completeness	13.16	2.22	10.94	5	38
06Oct06	ER Da...	Hospi...	Discharge Disposition	Percent Mapped to Known Values	13.16	2.22	10.94	5	38

Data Quality



Data Quality Frequencies

Geography	Rank	03Oct06	04Oct06	05Oct06	06Oct06	07Oct06	08Oct06	09Oct06	10Oct06
Hospital 01	1	FEVER (5)	CHEST PAIN...	ABD PAIN (5)	FEVER (5)	CHEST PAIN...	ABD PAIN (5)	FEVER (5)	CHEST PAIN...
Hospital 01	2	MVC (4)	ABD PAIN (5)	FEVER (5)	MVC (4)	ABD PAIN (5)	FEVER (5)	MVC (4)	ABD PAIN (5)
Hospital 01	3	SORE THRO...	FEVER (5)	BACK PAIN (...	SORE THRO...	FEVER (5)	BACK PAIN (...	SORE THRO...	FEVER (5)
Hospital 01	4	BACK PAIN (...	MVC (3)	CHEST PAIN...	BACK PAIN (...	MVC (3)	CHEST PAIN...	BACK PAIN (...	MVC (3)
Hospital 01	5	CONSTIPAT...	486-PNEUM...	780.2-SYNC...	CONSTIPAT...	486-PNEUM...	780.2-SYNC...	CONSTIPAT...	486-PNEUM...
Hospital 01	6	COUGH (2)	ABDOMINAL...	786.50-CHE...	COUGH (2)	ABDOMINAL...	786.50-CHE...	COUGH (2)	ABDOMINAL...
Hospital 01	7	CP (2)	BACK PAIN (...	HEMATURIA...	CP (2)	BACK PAIN (...	HEMATURIA...	CP (2)	BACK PAIN (...
Hospital 01	8	DIFF BREAT...	BLOOD IN U...	MVC (2)	DIFF BREAT...	BLOOD IN U...	MVC (2)	DIFF BREAT...	BLOOD IN U...
Hospital 01	9	MIGRAINE (2)	COUGH (2)	RT ANKLE I...	MIGRAINE (2)	COUGH (2)	RT ANKLE I...	MIGRAINE (2)	COUGH (2)
Hospital 01	10	SOB (2)	LOWER BAC...	SOB (2)	SOB (2)	LOWER BAC...	SOB (2)	SOB (2)	LOWER BAC...
Hospital 02	1	ABD PAIN (6)	ABD PAIN (8)	FEVER (6)	ABD PAIN (6)	ABD PAIN (8)	FEVER (6)	ABD PAIN (6)	ABD PAIN (8)

- **Frequencies will allow you to choose a text-based parameter and view the top 10 more common results.**
- **In a non-simulated version of ESSENCE, you will also be able to view the Data Quality – Hospital Status and Data Quality – Data Status pages to get information on data availability.**
- **This walkthrough is now complete.**

ESSENCE
Alerting Algorithms

Content was developed for and funded by the Centers for Disease Control and Prevention (CDC) for training purposes. The findings and conclusions in this presentation are those of the authors and do not necessarily represent the views of CDC.

ESSENCE Training Workshop

Statistical Alerting Algorithms

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

Content

- Overview
- Back-End vs. On-The-Fly
- Temporal (Single time series alerting)
  - Linear Regression
  - Exponential Weighted Moving Average (EWMA)
  - Regression / EWMA / Poisson Switch
  - Classical EARS methods C1 / C2 / C3
- Spatial Cluster Detection
- Time of Arrival: syndromic temporal clusters
- Summary Alerts: to control alert rate from many parallel streams
- Term-based: non-syndromic Alerting of Anomalous Chief Complaint Terms

Overview

- The purpose of the ESSENCE algorithms are to direct the attention of the users to data features that merit further investigation
- Algorithms in ESSENCE are not intended to identify outbreaks without supporting evidence.
- Algorithms in ESSENCE monitor for unusually high counts, not low counts (one-sided tests).
- Algorithms are designed to execute, produce prompt results in normal ESSENCE computing environments (not on supercomputers or very large clusters).

Overview

Major Types of Algorithms in ESSENCE include:

- Temporal
- Spatial
- Time of Arrival
- Summary
- Non-syndromic term-based alerting
- Fusion of multiple evidence types

Overview

Purpose:

- Temporal
  - Detect anomalous increases in cases over time (daily, weekly)
- Spatial
  - Detect geographic case clusters anomalous relative to a sliding baseline spatial distribution
- Time of Arrival
  - Detect temporal clusters of syndromic visits with similar arrival times (hourly)
- Summary
  - Provide alerts across numerous data streams adjusted for multiple testing
- Term-based Alerts (currently not in NSSP)
  - Find individual and unexpected terms in recent chief complaints that are anomalous relative to a baseline set
- Fusion: Bayesian Networks designed to emulate epidemiologist reactions to alerts across multiple syndromic/diagnostic data sources (currently only for DoD)

Back-End vs. On-The-Fly

In ESSENCE, the Alert List and myAlert pages are computed by algorithms running on a set schedule on back-end compute servers.

Time series graphs are color-coded red and yellow based on on-the-fly runs of the temporal detection algorithm chosen by the user.

This means that the alert list results can get out of sync with the time series results if newer data has been processed since the last time the back-end detection process has ran.

Temporal

Linear Regression

- Accounts for:
  - Linear Trend (seasonality)
  - Day-of-Week effects
  - Holiday effects
  - Day after Holiday effects

- 28 Day Baseline
- 2 Day Guard Band
- Outlier Removal
- Zero Filtration (avoids bias from data dropouts)
- Threshold p-values: .01 = Red, .05 = Yellow

Temporal

Exponentially Weighted Moving Average (EWMA)

- Performed at .9 and .4 smoothing coefficients (influence of recent past data)
- 28 Day Baseline
- 2 Day Guard Band
- Outlier Removal
- Zero Filtration
- Threshold p-values: .01 = Red, .05 = Yellow

Temporal

Switch Detector – Regression / EWMA / Poisson

- Performs Regression
- If baseline data pass goodness-of-fit test, Regression results used, else…
- Perform EWMA
- If there is not enough data in the baseline
- Perform Poisson
- 28 Day Baseline
- 2 Day Guard Band
- Outlier Removal
- Zero Filtration
- Threshold p-values: .01 = Red, .05 = Yellow

Temporal

EARS C1 / C2 / C3

- CDC Early Aberration Reporting System (EARS) Algorithms

Conventional settings:

- 7 Day Baseline
- No Guard Band
- No Outlier Removal
- No Zero Filtration
- Threshold p-values: 2 = Red, 1.5 = Yellow

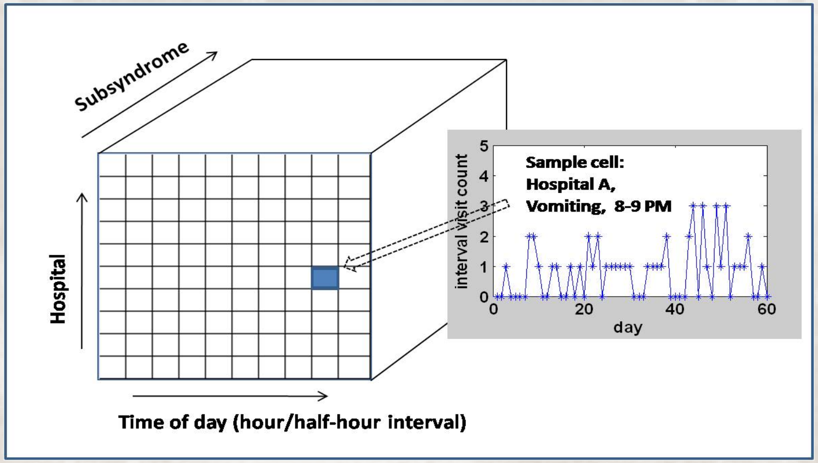
Spatial Cluster Detection

- Java-based Cluster Analysis based on methods in SaTScan software
- Zip Code based clusters
- 28 Day Baseline
- 2 Day Guard Band
- Test statistic: Kulldorff's Poisson log likelihood ratio
- Monte Carlo trials used to determine p-value (accelerated for rapid output)
- Threshold p-values: .01 = Red, .05 = Yellow

Time of Arrival

Finding clusters of visits linked by syndrome at similar times

- 60 Day Baseline
- Uses day of the week
- Inspection time blocks:
  - 60 minute on the hour
  - 30 minute
  - 60 minute on the half hour



- Performed by Hospital / Subsyndrome (special subset)
- Minimum 3 cases required to alert (may be increased by subsyndrome)
- Threshold p-value: 10^{-4} (0.0001)

Summary

Summary

- Used on Summary Alert List to derive a single resultant significance value from many parallel data streams.
- All data streams with p-values below the resultant value are considered to alert.
- To control alerting purely due to multiple testing.
- Uses a False Discovery Rate (FDR) based method.

Effect: alerts for

- a single alert of very high significance, or
- multiple alerts of joint relative significance

Summary

An example of the how the FDR detectors work is shown below. The algorithm starts by sorting all the input p-values. It then creates a multiplication factor based on the number of p-values (N) and the position in the sorted array (i). After you multiply the input p-value with the multiplier, you can take the minimum p-value and that becomes the summary alert p-value. The FDR-Major uses a modification that checks the input p-values and if at least half alerting, the input p-values are cut in half, and the FDR algorithm runs on the first half of the sorted input p-values.

P-Values (Sorted)	.0075	.0125	.0260	.4526	.7512
Multiplier (N/i)	5/1	5/2	5/3	5/4	5/5
FDR	.0375	**.0313**	.0433	.5658	.7512

P-Values (Sorted)	.0075	.0125	.0260	.4526	.7512
P-Values (Cut in Half)	.0075	.0125	.0260
Multiplier (N/i)	3/1	3/2	3/3
FDR Major	.0225	**.0188**	.0260

Because more than half are alerting, proceed with only the first half of the input P-Values

Red: .01

Yellow: .05

Word Alerts

Word Alerts

- Investigates frequency of individual words in text fields (like chief complaints) relative to pooled terms in 1-month baseline
- Uses Fisher's Exact Test
- For larger counts, uses chi-square test
- 30 Day Baseline
- 7 Day Guard Band
- Pvalue: 10^{-5} (0.00001)
- Not currently in NSSP

ESSENCE
Alerting Algorithms
Additional Reference Material

Content was developed for and funded by the Centers for Disease Control and Prevention (CDC) for training purposes. The findings and conclusions in this presentation are those of the authors and do not necessarily represent the views of CDC.

Center for Surveillance, Epidemiology, and Laboratory Services
Division of Health Informatics and Surveillance

Explanatory Overview of ESSENCE Alerting Algorithms

The following principles were written to clarify the use of univariate temporal algorithms in ESSENCE but apply to all of the methods described below:

General considerations:

1. These methods are not intended to positively identify outbreaks without supporting evidence. Their purpose is to direct the attention of a limited monitoring staff with increasingly complex data streams to data features that merit further investigation. They have also been useful for corroboration of clinical suspicions, rumor control, tracking of known or suspected outbreaks, monitoring of special events and health effects of severe weather, and other locally important aspects of situational awareness. Successful users value these methods more for the latter purposes and do not base public health responses solely on algorithm alerts.

2. All of these algorithms are one-sided tests that monitor only for unusually high counts, not low ones. Low counts could result from an emergency situation because data reporting could be interrupted, but there are many more common reasons for low counts (such as unscheduled closings or system problems), so the algorithms do not test for abnormally low counts.

3. In addition to data- and disease-specific considerations below, algorithm selection was also driven by system considerations. Users need to monitor many types of data rapidly. External covariates such as climate data or clinic schedules are not available for prompt analysis. Many methods in the literature, armed with substantial retrospective data of a certain type, depend on analysis of substantial history. Day-to-day users, often with only a small fraction of time available for monitoring, will not wait several minutes for each query. In the absence of data history and data-specific analysis time for each stream, ESSENCE methods have been adapted from the literature and engineered to system requirements.

4. If the time series monitored by algorithms represent many combinations of clinical groupings, age groups, and geographic regions, excessive alerting may occur simply because of the number of tests applied. The Summary Alert method was implemented to limit such excessive alerting. This method is based control of the *false discovery rate*, or the expected ratio of false alerts to the total alert count, and its statistical implementation in ESSENCE is detailed in the Summary Alert section below.

   Beyond analytic methods to control alerting, default alert lists should be limited to results from those time series of concern to the user, either by system design or by active specification by the user. For example, one method of reducing the default alert list is to restrict algorithms to all-age time series groupings. Depending on the scope of the user's responsibility, the alert list may also be restricted according to both epidemiological interest and the resources available for investigation. For example, a monitor of a national-level system with algorithms applied to many facilities may be interested only in alerts with at least 5-10 cases. In circumstances of heightened concern, these restrictions can be relaxed, or the user can use ESSENCE advanced querying methods to apply algorithms to age groups and/or subsyndromes.

The default temporal algorithm is an automated selection between data modeling (adaptive multiple regression) and control-chart-based (adaptive exponentially weighted moving average (EWMA)) algorithms, resorting to a simplistic (Poisson) method if only a few days of recent data are available. The primary regression and EWMA methods are discussed first separately.

Each description below gives a method category, purposes of the method, a brief technical description, key benefits, limitations, and literature sources.

Alerting Methods Applied to Single Time Series

1. Algorithm: Linear Regression

Categorization: Adaptive Multiple Regression Model

Purposes: This model is an adaptive regression model applied to remove the systematic behavior often seen in time series of daily, syndromic, clinical visit counts and in other surveillance data. The reason for removing these common effects is to avoid bias in identifying unusual behavior. For example, there is a customary jump in visits on Mondays because many clinics resume normal hours, and this expected jump should not automatically increase the possibility of an alarm. Similarly, alarms should be possible on weekends even though visit counts drop off from weekday levels.

Technical Details: This adaptive, multiple, least-squares regression algorithm contains terms to account for linear trends, day-of-week effects, and holidays. Multipliers for these terms are calculated using 4 weeks of recent counts as a training period. This training period is separated from the date of the test data by a 2-day buffer intended to keep early outbreak effects from contaminating the training. Extreme data values in the training period are reduced to reasonable values in order to avoid inappropriate predictions. This outlier correction for model inference avoids loss of sensitivity in the weeks after either data problems or true outbreaks.
The regression multipliers are recomputed each day for calculation of a predicted count based on the expected data trends. The algorithm then subtracts this prediction from the observed visit count, scales the excess by the standard error of regression, and applies a statistical hypothesis test to determine whether to signal an alert. The test is a Student's t-distribution at significance levels of 1% for red alerts and 5% for yellow alerts, with the number of degrees of freedom determined by the number of regression covariates and the baseline length.

Benefits: The main benefit is avoiding alerting bias resulting from expected data trends. The length for the training baseline is critical. Based on performance comparisons among multiple baseline lengths, it was chosen to be short and recent enough to capture seasonal time series behavior but long enough to smooth out daily fluctuations. Separate multipliers are updated so that a data source with regular but unusual patterns such as high weekend counts will be modeled correctly. While a better fit may often be obtained with a more complex model for a given data stream with a certain syndromic filter for a certain subregion and analysis of sufficient data history, the current regression approach is relatively robust across recent ESSENCE time series.

Limitations: If this algorithm is applied to a data series without the baseline weekly and seasonal behavior, the model will not explain the data well, and the detection sensitivity and specificity will be decreased. The automated switch in the default method is applied for this reason. There is no claim of optimal modeling for a given time series.

Sources:

1. Brillman JC, Burr T, Forslund D, Joyce E, Picard R and Umland E. Modeling emergency department visit patterns for infectious disease complaints: results and application to disease surveillance, BMC Medical Informatics and Decision Making 2005, 5:4, pp 1-14 http://www.biomedcentral.com/content/pdf/1472-6947-5-4.pdf.
2. Burkom, H.S., Development, Adaptation, and Assessment of Alerting Algorithms for Biosurveillance, Johns Hopkins APL Technical Digest 24 (2007), 4: 335-342

2. Algorithm: Adaptive Exponentially Weighted Moving Average (EWMA)

Categorization: Adaptive Control Chart

Purposes: This algorithm is appropriate for daily counts that do not have the characteristic features modeled in the regression algorithm. It is more applicable for Emergency Department data from certain hospital groups and for time series with small counts (daily average below 10) because of the limited case definition or chosen geographic region.

Technical Details: This algorithm compares a weighted average of the most recent visit counts to a baseline expectation. For the weighted average to be tested, an exponential weighting gives the most influence to the most recent observations. Two weightings are applied: the first gives negligible weight to observations over 3 days old and is designed to detect sudden events where most outbreak cases affect data within a few days. The second weighting distributes influence further over the past week for sensitivity to more gradual outbreaks.

The monitored weighted averages are the S_k given by:

$$S_k = \omega S_{k-1} + (1-\omega) X_k,$$

for a constant smoothing coefficient ω, with $0 < \omega < 1$ and X_k as the successive data counts, with $X_0 = 0$ and S_0 = half the alerting threshold for prompt sensitivity. (Occasionally a useful starting value for X_0 is known, but restarts may occur for many reasons, so the conservative initialization to 0 is used.) For separate monitoring of sudden and gradual events, smoothing coefficients $\omega = 0.9$ and 0.4 are used.

For both weighted averages, the 4-week baseline mean is subtracted, with a 2-day buffer period to separate the baseline from the counts being tested. The rationale for the baseline length was the same as described above for the regression method above. The test statistic is then $(S_k - \mu_k) / \sigma_k$, where μ_k , σ_k are baseline mean, standard deviation. As in the regression method, the hypothesis applied to determine alerting is a Student's t distribution at significance levels of 1% for red alerts and 5% for yellow alerts. The number of degrees of freedom is the baseline length + 1.

This algorithm is designed for any series that does not fit the characteristic trends, so safeguards are included for rapid adjustment to and recovery from data dropouts and catch-ups and for avoiding excessive alerts when counts are sparse.

Benefits: This method gives sensitivity to both sudden and gradual outbreaks and has demonstrated prompt alerting capability. It is less susceptible than the EARS methods C1, C2, and C3 to trends and to day-of-week effects. The added recovery features handle common problems in the data acquisition chain. Alerting is indirectly adjusted for the data distribution via the standardized residual test statistic, which provides a safeguard against excessive alerting when counts are small.

Limitations: This algorithm applied to pure daily counts does not control for expected trends or cyclic effects as in the regression method.

Sources:

1. Ryan TP. Statistical Methods for Quality Improvement. New York: John Wiley & Sons: New York, 1989
2. EWMA-Shewhart charts in Morton AP, Whitby M, McLaws M-L, Dobson A, McElwain S, Looke D, Stackelroth J, Sartor A; The application of statistical process control charts to the detection and monitoring of hospital-acquired infections; J Qual Clin Prac 2001; 21:112-117.

3. Algorithm: Poisson/Regression/EWMA (default)

Categorization: Automated switch between data model and control chart

Purpose: Many researchers and developers have applied complex statistical models to surveillance data for prediction and detection. However, the predictive capability of a model varies according to the specific data stream and how it is filtered and aggregated. This capability may also be affected by data behavior changes that result from seasonal variations, population shifts, and changes in the informatics. To account for such day-to-day changes, ESSENCE automatically monitors its predictive capability of its regression model each day. When this test fails, indicating that the model is not helpful for explaining the data, the system switches to the EWMA adaptation described above. The result is that the regression model is usually applied for the common respiratory and gastrointestinal syndrome classifications applied to county-level data, but EWMA is more commonly applied to rare syndrome data.

For situations where less than a week of recent baseline data exists, a simple Poisson detector is applied. Such situations include new start-ups and more common restarts after long (several-week) intervals of missing data.

Technical Details:
Details for the separate regression and EWMA methods are given in the preceding pages. The adjusted R^2 coefficient for the regression is tested each day. This coefficient does not give the quality of regression but is employed here specifically as a measure of daily predictive capability using an empirically derived threshold criterion. When the data pass this test, the model is assumed to have explanatory value, and the regression algorithm is applied. When the data fail this test, the EWMA algorithm is used.

The Poisson distribution test is applied when less than a week (3-6 days) of recent data is available. A Poisson distribution is assumed with mean and variance equal to the mean of the recent counts. An alert is issued if the current count exceeds this mean and if its probability is less than 1% (red alert) or 5% (yellow alert) according to the Poisson assumption.

For additional features engineered to meet the needs and requests of epidemiologist users, see the reference below.

Benefits: This algorithm is the default because it is designed to avoid mismatching the method to the data. The regression model accounts for the expected data trends when they are seen in the baseline. When they are absent because of the case definition used to filter the data, because of the size of the monitored region, or because of data problems, alerting is based on the EWMA algorithm.

Limitations: The goodness-of-fit test occasionally misclassifies the data. The test is set to err toward the more conservative EWMA to avoid mis-fitting the data model.

Sources:

Burkom HS, Elbert Y, Magruder SF, Najmi AH, Peter W, Thompson MW. Developments in the roles, features, and evaluation of alerting algorithms for disease outbreak monitoring. Johns Hopkins APL Technical Digest 2008;27:313.

4. Algorithms: C1, C2, and C3

Categorization: Adaptive Control Chart

Purpose: To purpose is to detect general data aberrations. Algorithms C1, C2, and C3 of the Early Aberration Reporting System (EARS) developed at the Centers for Disease Control and Prevention are used in many U.S. states and in numerous foreign countries. They are included in the ESSENCE suite because of their wide application. While they lack many of the features described above, their simplicity has both benefits and limitations.

Technical Details: The C1 algorithm subtracts the daily count from the mean of a moving baseline ending the previous day. In effect, it then divides this difference by the standard deviation of counts in that baseline. If the result exceeds 3, indicating an increase above the mean of more than 3 standard deviations, an alert is issued.

The C2 algorithm does the same calculation but imposes a 2-day buffer between the test day and the baseline.

The C3 algorithm is a more sensitive version of C2 that adds the values from the 2 previous days if they do not exceed the threshold. All three algorithms use the same criterion of an increase of at least 3 baseline standard deviations above the sliding baseline mean.

An important implementation detail is that ESSENCE does not use the standard 7-day baseline because substantial experience has shown that for many time series, such a short baseline gives an unstable statistic that can lead to a loss of confidence in the results. The implemented baseline is 28 days as in the EWMA and regression methods. There are no other changes to the standard EARS methods, including retention of the flat 3-standard-deviation threshold regardless of the data stream.

Benefits: The methods are easy to understand and widely known.

Limitations: Like the EWMA, the methods take no account of systematic data behavior such as day-of-week effects or seasonal trends. C3 is the only one of these methods with sensitivity to gradual outbreak effects, but it is known to produce high alarm rates. For all three methods, threshold data values for alerting may fluctuate noticeably from day to day.

Sources:

1. Hutwagner LC, Maloney EK, Bean NH, Slutsker L, Martin SM. Using laboratory-based surveillance data for prevention: an algorithm for detecting Salmonella outbreaks. Emerg Infect Dis 1997; 3:395–400
2. Tokars JI, Burkom HS, Xing J, English R, Bloom S, Cox K, and Pavlin JA, Enhancing Time-Series Detection Algorithms for Automated Biosurveillance, Emerg Infect Dis. 2009 Apr;15(4):533-9.

Epidemiologic investigation involves analyzing the geographic distribution of cases to determine if an outbreak is associated with a geographic region. Geographic information systems (GIS) are tools that allow spatial mapping of data. In ESSENCE systems, data visualization is performed with the geo-spatial analysis software, Geoserver. This GIS capability assists the user in determining if an anomaly in syndrome counts is localized, and it may aid in the identification of a point-source disease outbreak. GIS may also help in predicting the geographic extent of the affected population to expedite the correct allocation of public health resources. In addition to spatial mapping, ESSENCE uses spatial scan statistics to search for unexpected clustering of cases for each of several syndrome groups.

Spatial Cluster Determination

Category: Spatial Scan Statistics

Purpose: A problem with sophisticated temporal detectors is choosing the appropriate size and location of the collection region for time series counts. If this region is too small or mislocated, cases may be missed and the baseline data may not have enough structure, but if the region is too large, the scale and variability of the large-scale time series may reduce sensitivity by masking clusters of interest. We apply spatiotemporal scan statistics in an attempt to promptly localize public health problems. For ESSENCE, JHU/APL built and implemented a Java version of the SaTScan software of Martin Kulldorff originally developed for spatial surveillance of cancer and subsequently used and enhanced for many types of hotspot detection.

Technical Details: The null hypothesis is that the set of data subregions (often patient zip codes) in the recent time interval tested forms a random sample from an expected spatial distribution of cases. The expected distribution is not uniform over subregions but reflects a "customary" spatial case spread that reflects urban/suburban case ratios or other factors. ESSENCE implementation calculates the expected spatial distribution using recent case counts from a sliding baseline interval. In effect, the code is similar to a common application of SaTScan, the space-time permutation scan statistic, restricted to test cases from only the most recent time interval and assuming circular clusters.

As in SaTScan, the method calculates a test statistic for each candidate cluster. The test statistic in the ESSENCE implementation is Kulldorff's Poisson log likelihood ratio. The set of candidate clusters is generated by scanning over a set of cluster center locations, often taken as centroids of all zip codes in the dataset, and considering all circles within a maximum radius of each center, where the number of circles is limited by the number of data subregions within each radius. The maximum test statistic over these candidates is then tested for significance.

Statistical significance inference does not depend on a theoretical distribution but on repeated trials on simulated datasets randomly drawn using the baseline distribution. For each such trial, the algorithm uses the same scanning procedure to derive a trial maximum.

For assessing the significance of the maximum test statistic over all observed clusters, the ESSENCE code uses the Gumbel distribution method as published by Abrams, Kleinman and Kulldorff. The code collects 99 trial maxima, fits a Gumbel distribution to these values, and uses the fitted distribution to assign a p-value to the test statistics of clusters found in the original data. The observed cluster with the maximum test statistic is considered significant if its p-value is below a predetermined threshold, often set to 0.01. This threshold criterion can yield multiple significant clusters in a given run if more than one candidate cluster yields a test statistic whose p-value is below the threshold.

For each significant case cluster, the system shows the location, extent, and degree of significance using the GIS software.

Benefits: The ESSENCE Java implementation inherits features that have popularized SaTScan. Potential clusters of interest are localized without bias regarding the center or extent of the cluster as well as the spatial resolution of the data allows. As noted in Kulldorff, Heffernan, et al., the empirical significance testing with many repeated trials takes "into account the multiple testing stemming from the many potential cluster locations and sizes evaluated."

Limitations: The most important limitation, applicable also to SaTScan and to all other spatial or space-time cluster detection methods, is that the usefulness of the method strongly depends on the reliability of the expected spatial distribution. The use of census-based distributions, insurance eligibility lists, regression models, and other means have been used to derive the expected distribution. The method implemented in ESSENCE infers this distribution from recent data separated from the test date(s) by a 2-day buffer.

Evaluation of statistically significant clusters for epidemiological significance is a nontrivial task which may be exacerbated if the number of significant clusters is misleading or excessive because the expected distribution is unrepresentative or because investigation resources are insufficient.

The use of this popular approach has been criticized for prospective use; see Correa et al. The ESSENCE implementation lacks the controls applied in the prospective version of SaTScan attempting to manage cluster rates for multiple successive days. The ESSENCE implementation does support elliptical cluster shapes, simultaneous clustering of multiple data sources, or test statistics other than the Poisson log likelihood ratio, and the user with a sufficiently detailed dataset and an application that requires extended SaTScan features should be aware of these limitations.

Sources:

Kulldorff M. A spatial scan statistic. Communications in Statistics–Theory and Methods. 1999;26:1481-1496.

Kulldorff M, Heffernan R, Hartman J, Assunção R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. PLoS Medicine, 2005; 2:216-224.

Correa, T.R., R.M. Assuncao, and M.A. Costa. 2015. A critical look at prospective surveillance using a scan statistic. Stat. Med. 34(7): 1081–1093. doi:10.1002/sim.6400.

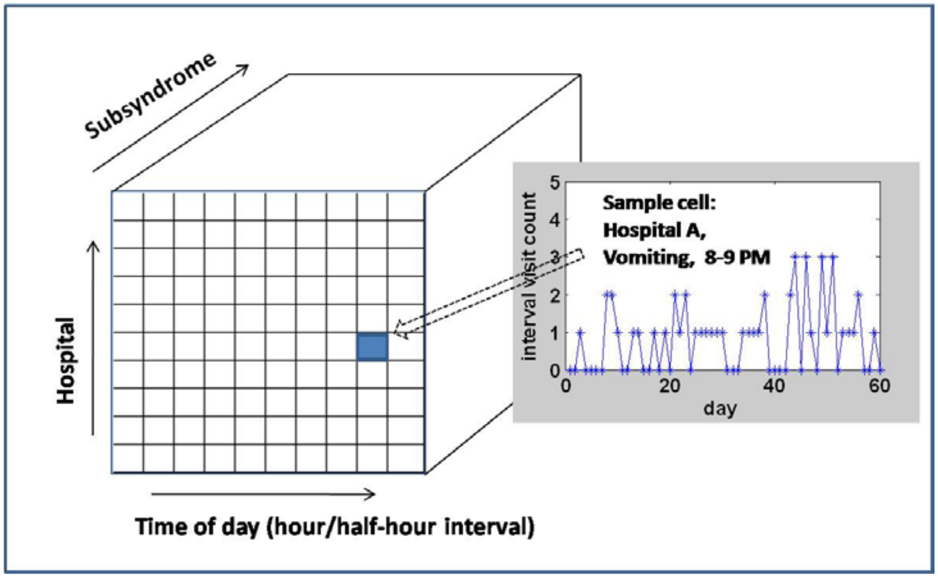
Abrams A., Kleinman K., Kulldorff M. Gumbel based p-value approximations for spatial scan statistics. International Journal of Health Geographics 2010 9:61, DOI: 10.1186/1476-072X-9-61

Time-of-Arrival Cluster Determination

Categorization: Multiple Automated Hypothesis tests

Purpose: This algorithmic approach was implemented to find and display unusual clusters of syndromically related emergency department visits by patients arriving for care within a short time interval.

Technical Details: Patient visit counts are tabulated by cells, with one cell for each hospital/time-interval/sub-syndrome combination. See Figure 1.



- For the visit counts in each cell, a Poisson or negative binomial test is chosen using the last 60 days of visit counts for that cell. The Poisson distribution is used unless the count variance exceeds the mean by a factor of 1.1 or greater, and then the time series is considered overdispersed. This situation occurs for relatively few cells, generally corresponding to the more common (sub) syndromes for the largest hospitals at the busiest times when most alerts would be generated. For this situation, a negative binomial distribution is assumed.
- Once the distribution is chosen, parameters for each cell are calculated from the 60-day baseline. For each cell, an alert is then flagged if the current count exceeds the upper limit threshold for the chosen distribution based on a preselected p-value.
- Based on empirical results using 12 years of data from 134 hospital EDs from a large state with labeled events, a threshold p-value of p* = 10-4 (0.0001) was chosen.
- Time intervals for the cells are 30 min., 60 min. beginning on the hour, and 60 min. beginning on the half hour, again a result of empirical testing.

- Practical overrides are implemented based on observed cell counts. At least three observed cases are required for an alert. This minimum may be increased for more common syndromes. Mandatory alerts may also be implemented for certain subsyndrome/count combinations, such as subsyndromes for severe illness, regardless of the hypothesis test.

Benefits: In validation testing to monitor visit clusters for 51 subsyndromes for 134 hospitals at the time intervals above with the chosen p-value threshold, alert rates were consistently manageable and found all known clusters from a small historical collection of events except for two groups of 3-4 visits at very busy times. The alert burden was still manageable at the county level when anomalous clusters for all hospitals within each county were combined.

The simplicity of this approach allows multiple daily runs and adaptation to new improvised subsyndromes with rapid system response without impact on routine processing.

Limitations: The hypothesis tests include no direct modeling of seasonality or other systematic data behavior. They were implemented to enable county-level processing, and validation was conducted on a 12-year historical dataset from one state. Expanding the computational load to include much larger sets of hospitals or syndrome groups with limited investigation capability may require recalibration (p-value threshold, minimum alert counts) or an alternate approach to retain sensitivity with manageable alerting.

Sources: H Burkom, L Ramac-Thomas, R Arvizu, C Lee, W Loschen, R Wojcik, and A Kite-Powell, A collaboration to enhance detection of disease outbreaks clustered by time of patient arrival. Emerging Health Threats Journal 2011, 4:s65. doi: 10.3134/ehtj.10.065.

Summary Alert Algorithm

Categorization: False Discovery Rate processing of multiple alerts

Purpose: The *parallel monitoring problem* is the monitoring of many parallel time series representing different physical locations, such as counties or treatment facilities, possibly stratified by other covariates such as syndrome type or age group. The purpose of the Summary Alert Algorithm is to maintain sensitivity while limiting the number of alerts that arise from testing the numerous resulting time series.

Multiple testing can lead to uncontrolled alert rates as the number of data streams increases. For example, suppose that a hypothesis test is conducted on a time series of daily diagnoses of influenza-like illness. In a one-sided test, this test results in a statistic whose value in some distribution yields a probability p that the current count is as large as observed. For a desired Type I error probability of α, the probability is then (1- α) that an alert will not occur in the distribution assumed for background data. Thus, for the parallel monitoring problem of interest here, if such tests are applied to N independent data streams, the probability that no background alerts occur is $(1-\alpha)^N$, which decreases quickly for practical error rates α. For a single-test error rate of $\alpha = 0.05$, for example, the probability of at least one background alert exceeds 0.5 if more than 13 independent tests are applied.

Technical Details:
For N tests, where N is the number of combinations of region, syndrome, age group, and any other covariates affecting the number of tests, let $P_{(1)},\ldots, P_{(N)}$ be the *p*-values sorted in ascending order, an ordering that puts the smallest and most significant p-value first. The Summary Alert method applies the Simes-Seeger-Eklund criterion to reject the combined null hypothesis of no anomaly for any series. The null hypothesis is rejected if for some j^*, $j^* = 1,..,N$, $P_{(j^*)} < j^*\alpha / N$. To interpret this condition, note that for the most significant p-value, an alert requires that $P_{(1)} < \alpha/N$, the strict Bonferroni bound. If α=0.01 and N=50, then the condition becomes $P_{(1)} < 0.0002$. For the least significant p-value, the condition is simply $P_{(N)} < \alpha$, highly unlikely for the weakest result.

If this condition is satisfied for any j*, then test results are considered alerts for all j < j*. The Summary Alert is implemented at two levels, FDR and FDR-Major. For the FDR level applied to N time series, the implementation is as above. For a more liberal option appropriate for certain syndromes or scenarios, FDR-Major applies the condition to two sets of N/2 time series.

Benefits: In defining the false discovery rate as the expected ratio of false alerts to the total alert count, Benjamini and Hochberg showed that the Simes-Seeger-Eklund criterion gives an overall error rate of α if the N time series tested are statistically independent. Overall, this criterion avoids the excess alerting resulting from using the nominal threshold α for all data streams and also avoids the loss of sensitivity from using only the Bonferroni bound α/N.

Limitations: If one of the p-values crosses the adjusted threshold, it is not obvious for epidemiological or other reasons which tests to consider anomalous. Most users have followed the natural procedure described by Simes to consider all p-values less than $P_{(j^*)}$ as individual alerts. Another limitation is that in general the time series are not statistically independent. For situations where dependence is known, Hommel recommended the condition $P_{(j)} < j \cdot \mathrm{i} / C \cdot N$, where $C = \Sigma\ 1/j$. In ESSENCE applications where many groups of time series may be requested and dependence can change, the above condition with C=1 is applied.

Sources:
Simes RJ. An improved Bonferroni procedure for multiple tests of significance. Biometrika 1986;73:751-754.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Statistical Society B 1995; 57:289-300.

Hommel G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika 1988;75:383-386.

ESSENCE
Chief Complaint Processor

Content was developed for and funded by the Centers for Disease Control and Prevention (CDC) for training purposes. The findings and conclusions in this presentation are those of the authors and do not necessarily represent the views of CDC.

ESSENCE Training Workshop

Chief Complaint Processor

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

Content

- High Level Overview
- Specific Capabilities
  - Weights
  - Rules
  - Attributes
  - Abbreviation Expansion
  - Special Abbreviations
  - Fuzzy Matching
  - Dictionary
  - Negation
  - Stop Words
- Configuration Options
- CCDD
- User Interface in ESSENCE

High Level Overview

- Chief Complaints can be any type of text string:
  - **1 word:** “fever”
  - **small number of words:** “shortness of breath”
  - **verbose text:** “patient was seen with a cough that had been persistent for 3 weeks along with additional head aches and chills”
  - **with abbreviation:** “sob”
  - **with negation:** “patient was not vomiting”
  - **with misspellings:** “patient was not vomiting”
  - **in first person:** “I am having chest pain”
  - **in other languages:** “estoy teniendo dolor en el pecho”
  - **or any combination of the above**

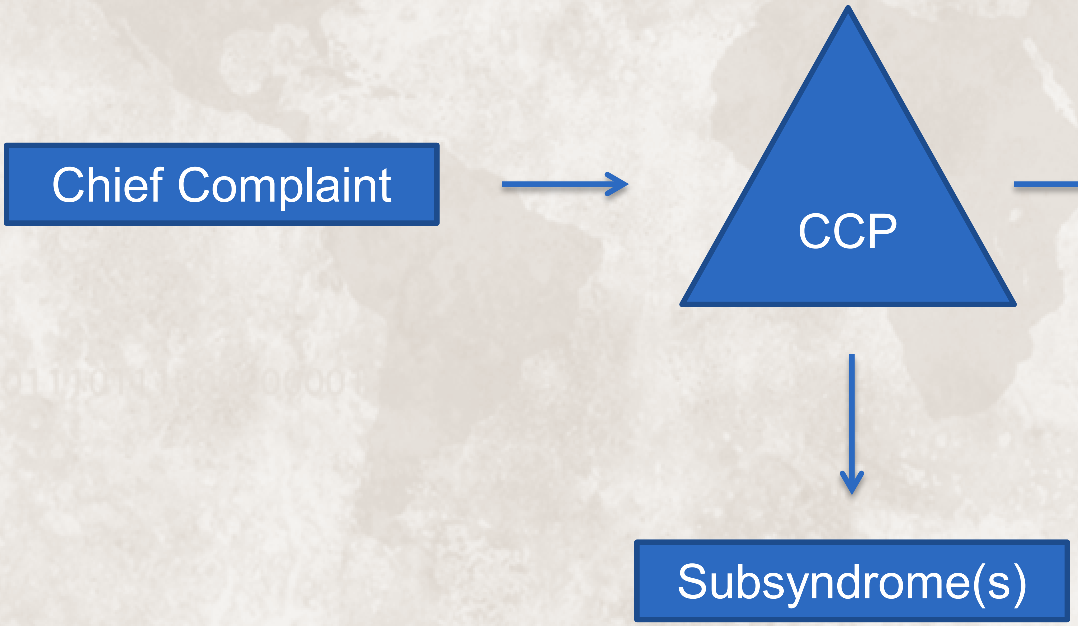
High Level Overview

- The ESSENCE Chief Complaint Processor (CCP) categorizes text into as many syndromes and subsyndromes as the text matches into.

- **Syndrome:** a group of associated symptoms
  - **Fever**
  - **GI**
  - **Respiratory**

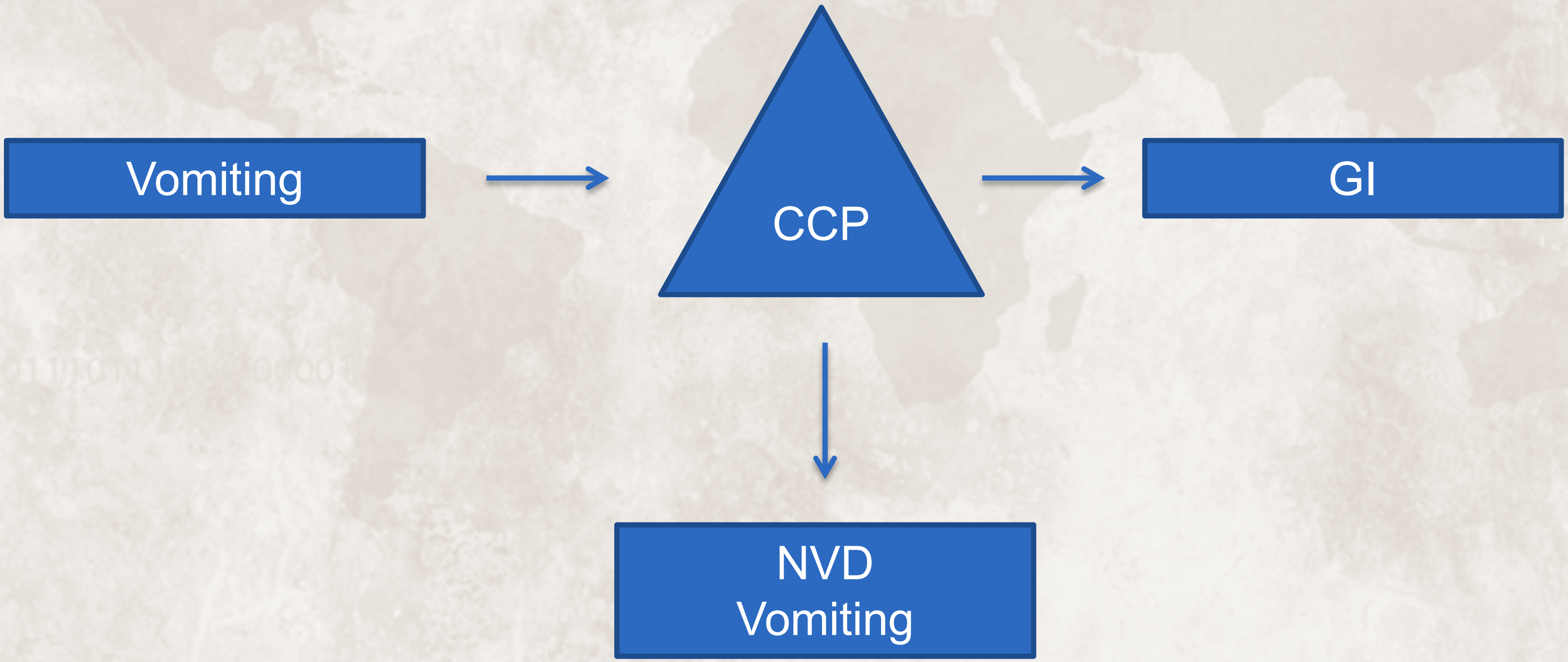
- **Subsyndrome:** a smaller, more specific group of associated symptoms
  - **Abdominal Pain**
  - **Difficulty Breathing**
  - **Diarrhea**

High Level Overview



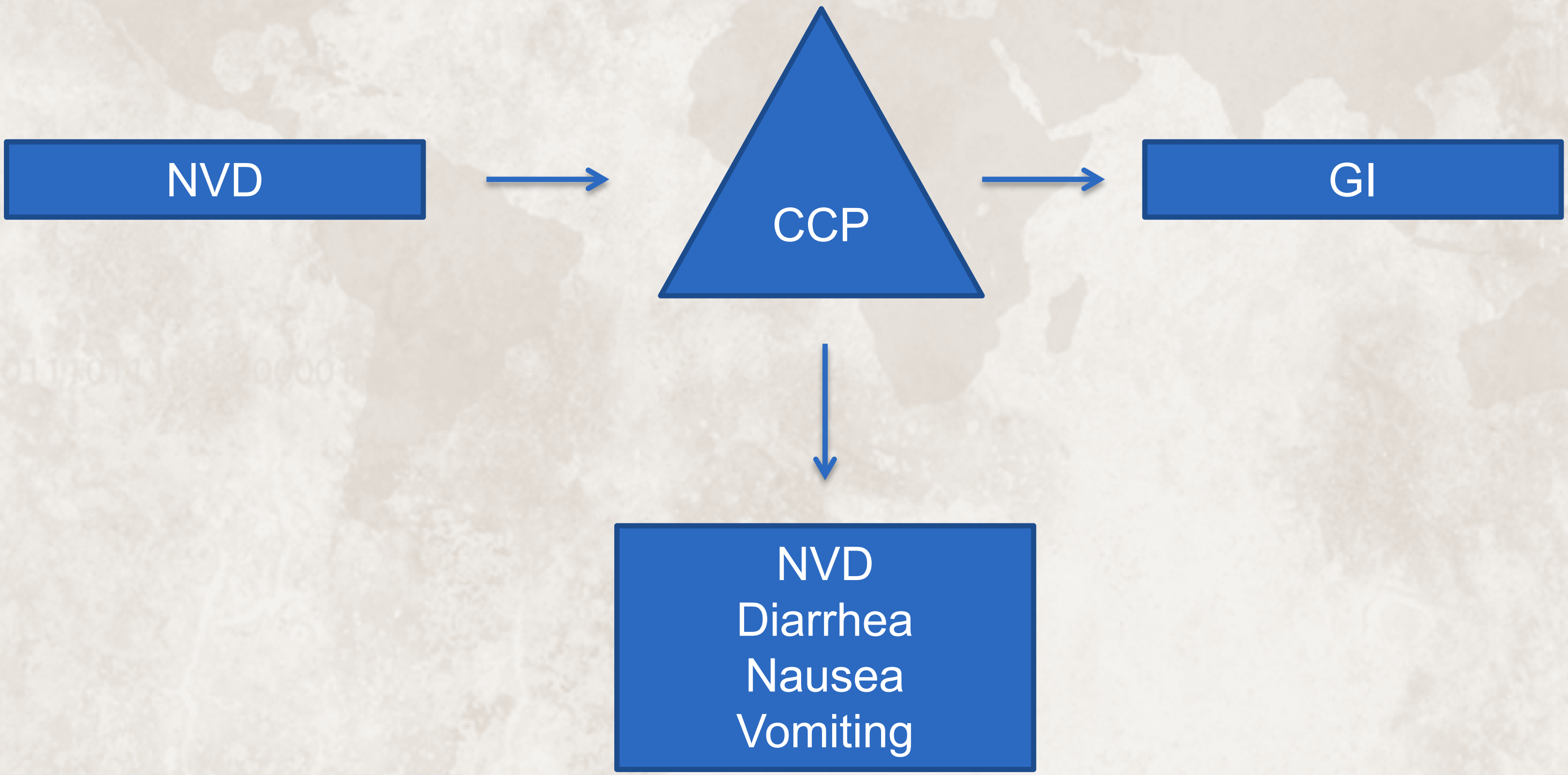
High Level Overview

Easy:



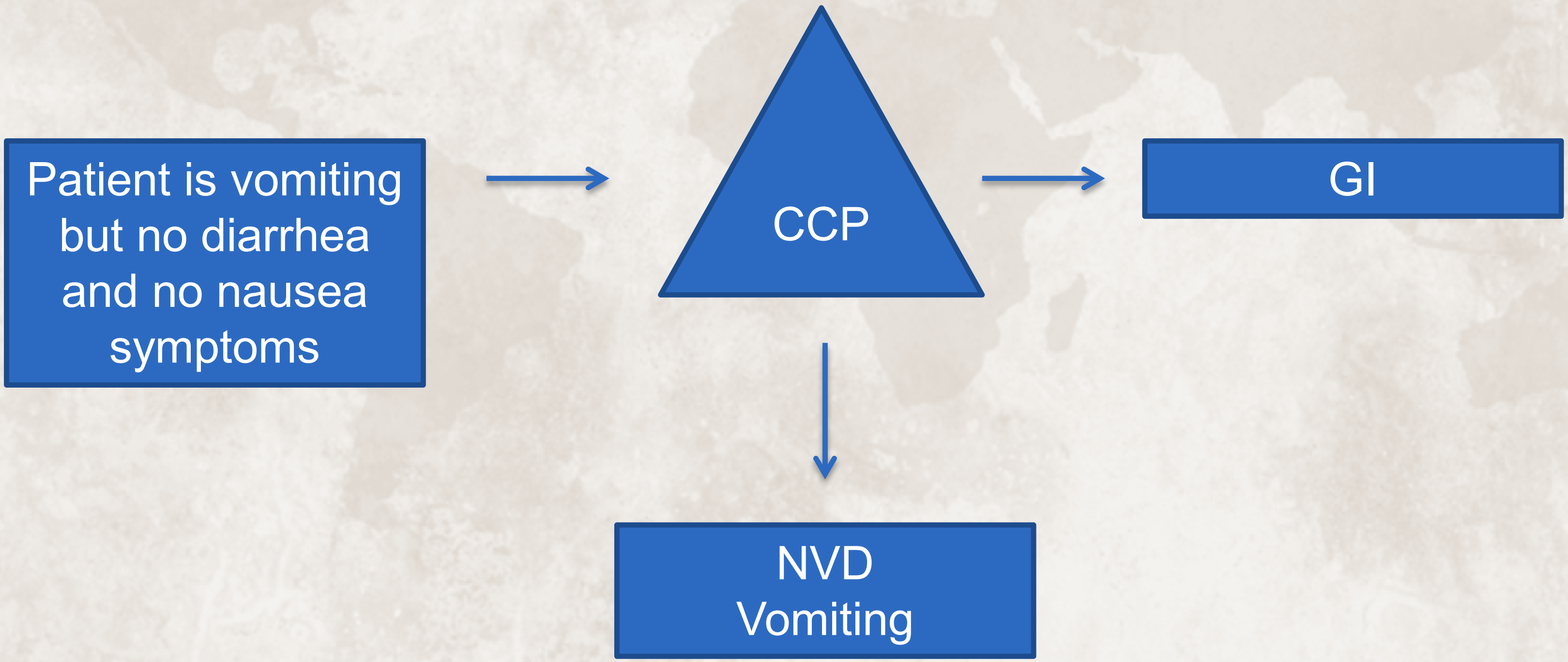
High Level Overview

Harder:



High Level Overview

Even Harder:



Specific Capabilities

Weights

- CCP uses a weighted keyword matching system
- 6 points required for a match
- Positive or Negative Numbers
- Wildcards allowed
- GIBleeding:

BELLY (4)	BELLY ACHE (-4)	BELLY PAIN (-4)
BLACK (2)	BLED (2)	BLEED (2)
BLEEDING (2)	BLOOD (2)	BLOOD PRESSURE (-2)
BLOOD SUGAR (-2)	BLOODY (2)	BOWEL (4)
DIARRHEA (4)	FECAL (4)	FECES (4)
GASTROINTESTINAL (4)	HEMATOCHEZIA (6)	HEMORRHAGE (2)
HEMORRHAGING (2)	INTESTINAL (4)	INTESTINE (4)
MELENA (6)	RECTAL (4)	RECTUM (4)
STOMACH (4)	STOMACH ACHE (-4)	STOMACH PAIN (-4)
STOOL (4)	TARRY (2)	TOILET PAPER (4)
VOIDED (4)		

Specific Capabilities

Rules

- CCP allows for Rules, Terms, or combinations to determine a subsyndrome or syndrome
- Rules are logical expressions of subsyndromes

Neuro = AlteredMentalStatus or Dizziness or Drowsiness or Encephalitis or (Headache and Fever) or ProjectileVomiting or Prostration or Seizure or SidedWeakness

ILI = Influenza or (Fever and (Cough or SoreThroat) and not NonILIFevers)

Specific Capabilities

Attributes

- CCP allows for attributes to be injected into the rules
- Injects information from the patient record to be used by the CCP

Resp = (Anthrax or Bronchitis or (ChestPain and **[Age<50]**) or Cough or Croup or DifficultyBreathing or Hemothorax or Hypoxia or Influenza or Legionnaires or LowerRespiratoryInfection or Pleurisy or Pneumonia or RespiratoryDistress or RespiratoryFailure or RespiratorySyncytialVirus or RibPain or ShortnessOfBreath or Wheezing) and not (GeneralExclusions or Cardiac or (ChestPain and Musculoskeletal) or Hyperventilation or Pneumothorax)

Specific Capabilities

Abbreviation Expansion

- Attempts to expand abbreviations
- Can only match a single abbreviation
- Abbreviations can have positive and negative requirements

NVD = NAUSEA VOMITING DIARRHEA
Positive Requirement: None
Negative Requirement: None

N = NAUSEA
Positive Requirement: None
Negative Requirement: '* D N *' OR '* N V *' OR '* N V D *' OR '* H1N1 *'

Specific Capabilities

Abbreviation Expansion

- Can get complicated…
- Abbreviation, Subsyndrome, Positive, Negative

AB, ABRASION, '* CORNEA*AB *' OR '*CONJ*AB *', none

AB, ABORTION, none, '* PAIN *' OR '* WOUND *' OR '* FEVER *' OR '* LAP *' OR '* LAPAROSCOPIC *' OR '* DISTEN*'

AB, ABDOMINAL, '* PAIN *' OR '* WOUND *' OR '* FEVER *' OR '* LAP *' OR '* LAPAROSCOPIC *' OR '* DISTEN*', none

AB, ABUSE, '* CHILD AB *', none

Specific Capabilities

Special Abbreviations

- Specifically converted during the CCP process, then have the ability to be put back when finished

1st, FIRST, false
2nd, SECOND, false
3rd, THIRD, false
4th, FOURTH, false
5th, FIFTH, false
6th, SIXTH, false
7th, SEVENTH, false
8th, EIGHTH, false
9th, NINTH, false
10th, TENTH, false
H1N1, HONENONE, true
#1H1N1, POUND_ONE_HONENONE, true
#2H1N1, POUND_TWO_HONENONE, true
#3H1N1, POUND_THREE_HONENONE, true
1H1N1, ONE_HONENONE, true
2H1N1, TWO_HONENONE, true
3H1N1, THREE_HONENONE, true
#1 H1N1, POUND_ONE_SP_HONENONE, true
#2 H1N1, POUND_TWO_SP_HONENONE, true
#3 H1N1, POUND_THREE_SP_HONENONE, true
1 H1N1, ONE_SP_HONENONE, true
2 H1N1, TWO_SP_HONENONE, true
3 H1N1, THREE_SP_HONENONE, true

Specific Capabilities

Fuzzy Matching

- Will attempt to match a word to a term if it is:
  - **1 letter inserts:**
  - chest = .chest | c.hest | ch.est | che.st | ches.t | chest.
  - **1 letter deletes:**
  - chest = hest | cest | chst | chet | ches
  - **1 letter substitutions:**
  - chest = .hest | c.est | ch.st | che.t | ches.
  - **1 letter inversion:**
  - chest = hcest | cehst | chset | chets

Specific Capabilities

Dictionary

- Terms that are in the dictionary, are NOT fuzzy matched
- Default ESSENCE implementation has 1855 dictionary terms

CRASH – Prevents fuzzy matching into RASH

HEAD – Prevents fuzzy matching into HEAT

A FEVER – Prevents fuzzy matching into Q FEVER

Specific Capabilities

Negation

- Two versions of Negation in the CCP
- "Original" and "Nebraska" mode
- "Nebraska" mode was built to handle chief complaints that were more like Triage Notes.

- Original = Negative then Term

no fever
not vomiting

DENIES
NEGATIVE
NO
NO EVIDENCE
NO EVIDENCE OF
NOT
NOT COMPLAINING OF
NOT COMPLAINING OF A
WITHOUT
WITHOUT MENTION
WITHOUT MENTION OF

Specific Capabilities

Negation

- Nebraska mode:
  - Negative then Term

no FEVER

  - Negative then 1 or 2 words then AND/OR then term

no cough, chills, or FEVER

  - Negative then 1 word then term then AND/OR

no cough, FEVER, or chills

  - If term supports reverse negation:
    - Term then Negative

FEVER denied

    - Term then 1 or 2 words then Negative

FEVER is denied

Specific Capabilities

Stop Words

- A stop word is a phrase that will be removed entirely from the input stream before processing.

AN
AND
CENTIMETER
DAY
DAYS
HOUR
HOURS
IN
METER
MONTH
MONTHS
ND
RD
TH
THE
WEEK
WEEKS

Configuration Options

- Which Pre-processors to turn on
  - Upper case
  - Punctuation
  - Abbreviation
  - Stop Words
- What attributes to include
  - Age
- Term Weight Threshold
  - 6
- Minimum Fuzzy Match Length
  - 5
- Negation Mode
  - Original / Nebraska

CCDD

- In addition to the Chief Complaint Processing into Syndromes and Subsyndromes, and additional text processing occurs on the CCDD field.

- CCDD is a concatenated field of the Chief Complaint (parsed) and the Discharge Diagnosis fields.

- Currently, there are 2 normal CCDD categories:
  - Foreign Travel
  - Visits of Interest

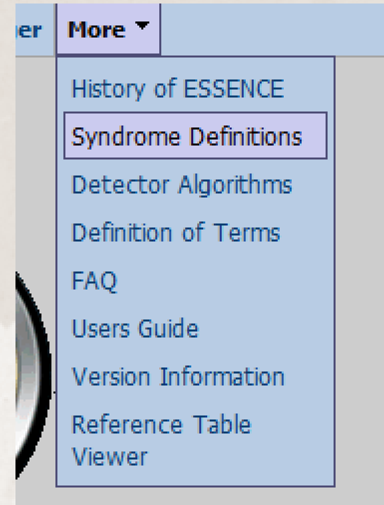
CCDD

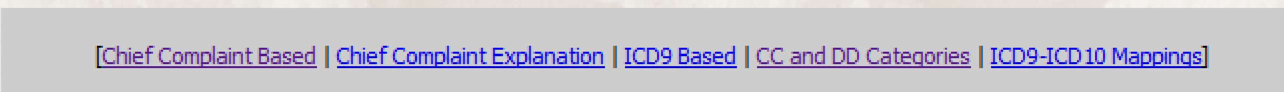
- CCDD Categories use SQL where clauses to find records that meet the criteria.
- For the most part, this is simple keyword matching.
- There are some wild-cards and some negation terms.
- The CCDD is wrapped in spaces to help find individual words.

- Examples:
  - ‘ ‘ + Foreign Travel + ‘ ‘ like ‘% chile %’ OR
  - (‘ ‘ + Foreign Travel + ‘ ‘ like ‘% china %’ AND NOT
  - ‘ ‘ + Foreign Travel + ‘ ‘ like ‘% hutch %’ AND NOT
  - ‘ ‘ + Foreign Travel + ‘ ‘ like ‘% cabinet %’) OR …

User Interface in ESSENCE

- Click on the More tab in ESSENCE
- Choose Syndrome Definitions





- The “Chief Complaint Based” option will describe the syndromes derived from the Chief Complaint using the CCP

User Interface in ESSENCE

- The Rules and/or Terms that a syndrome or subsyndrome is defined by can be viewed:

SYNDROMES > **NEURO**
Rules
Neuro = AlteredMentalStatus or Dizziness or Drowsiness or Encephalitis or (Headache and Fever) or ProjectileVomiting or Prostration or Seizure or SidedWeakness
Terms
N/A

SYNDROMES > NEURO > **DIZZINESS**

Rules

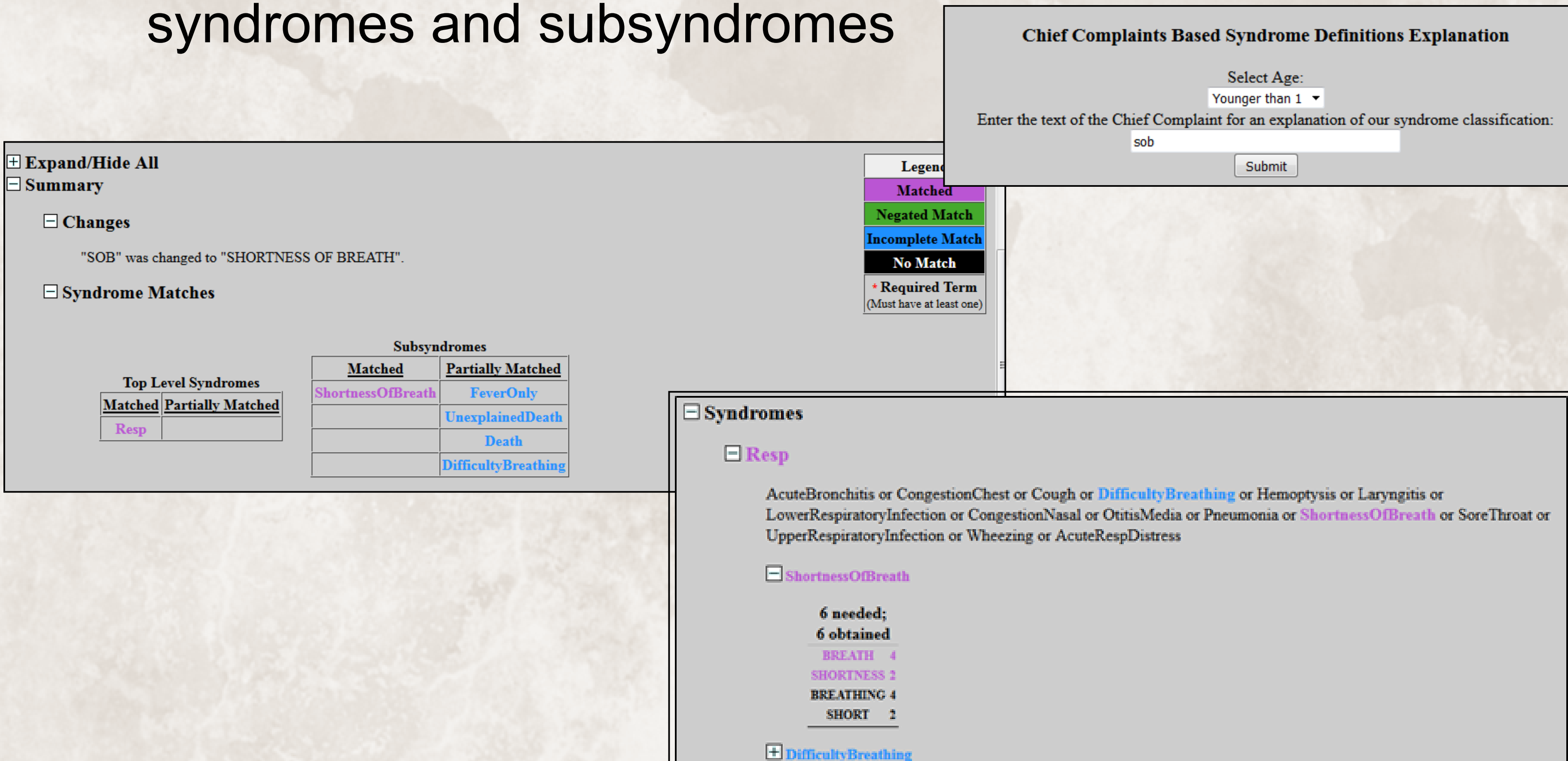
Dizziness = N/A

Terms

ABOUT TO FAINT (10)	ALMOST (2)	BALANCE (4)
DISEQUILIBRIUM (0)	DIZZINESS (10)	DIZZY (10)
EQUILIBRIUM (4)	FAINT (6)	FAINTED (4)
FEELING FAINT (10)	FEELS FAINT (10)	FEELS LIKE FAINTING (10)
FELT FAINT (10)	HEAD (2)	LACK (2)
LIGHTHEADED (10)	LIGHTHEADEDNESS (10)	LIKE (2)
LOSING (2)	LOSS (2)	NEAR (2)
NEARLY (2)	OFF (2)	PASS OUT (4)
PASSED OUT (4)	PASSING OUT (4)	ROOM (2)
SPINNING (4)	SYNCOPAL (4)	SYNCOPE (4)
VERTIGO (10)	WOOZIE (10)	WOOZINESS (10)
WOOZY (10)		

User Interface in ESSENCE

- The “Chief Complaint Explanation” page allows you to type in a chief complaint, and see how it will mapped into syndromes and subsyndromes



Questions? We appreciate your input.

Michael A. Coletta, MPH
Manager, National Syndromic Surveillance Program
CDC/CSELS/DHIS
mcoletta@cdc.gov